

ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

09-547-267-3
Query Match 44.7%; Score 686.5; DB 4; Length 303;
Best Local Similarity 48.7%; Pred. No. 1.6e-67;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
1 MAVGSKSPATSTLFDKTRSVMLYAWCHCDVDDQTLGFHADQPSQMPQRLQQ 60
12 IAQGSQFAAKLMPGPIREDTVMYAWCHHADVDIGQVNGSAPEAGGD--PQALGA 69
61 LEMKTRQAVAGSQMEE-----PAPAAFOEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTL 115
70 LEADTLAA-----LHEDGPMSPPPAALQVARRHDFPDLMPMDLIEGFAMDVADREYRSL 124
116 DDTLYCHVAGVGLMMAQIMGVNDNATLDRACDLGLAFOLTNIAARDVDDAQVGRCYL 175
125 DDVLEYSYHVAGVGVMMARVMGVQDDAVLDACDLGLAFOLTNIAARDVDDAAIGRCYL 184
176 PESWLEEGELTKANYAAPENQALSRAGRLVRAEPPYVSSMAGLAQLPLRSAAWATA 235
185 PADWLAEG--ATVEGVPSDALYSVIIRLDRAEPPYASARQGLPHLPPRCAWSIAAA 241
236 KQVYRKIGVKVEQAGKQAWDHRSQSTSTAEKLTLLLTASGQAVTSMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIQGGPEAYRQRISTSKAATIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

RESULT 13
3-09-920-923B-3
Sequence 3, Application US/09920923B
Patent No. 6671134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/12594)
CURRENT APPLICATION NUMBER: US/09/920-923B
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in version 3.1
SEQ ID NO 3
LENGTH: 303
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534

US-09-920-923B-3

Query Match 44.7%; Score 686.5; DB 4; Length 303;
Best Local Similarity 48.7%; Pred. No. 1.6e-67;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
QY 1 MAVGSKSPATSTLFDKTRSVMLYAWCHCDVDDQTLGFHADQPSQMPQRLQQ 60
DB 12 IAQGSQFAAKLMPGPIREDTVMYAWCHHADVDIGQVNGSAPEAGGD--PQALGA 69
QY 61 LEMKTRQAVAGSQMEE-----PAPAAFOEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTL 115
DB 70 LEADTLAA-----LHEDGPMSPPPAALQVARRHDFPDLMPMDLIEGFAMDVADREYRSL 124
QY 116 DDTLYCHVAGVGLMMAQIMGVNDNATLDRACDLGLAFOLTNIAARDVDDAQVGRCYL 175
DB 125 DDVLEYSYHVAGVGVMMARVMGVQDDAVLDACDLGLAFOLTNIAARDVDDAAIGRCYL 184
QY 176 PESWLEEGELTKANYAAPENQALSRAGRLVRAEPPYVSSMAGLAQLPLRSAAWATA 235
DB 185 PADWLAEG--ATVEGVPSDALYSVIIRLDRAEPPYASARQGLPHLPPRCAWSIAAA 241
QY 236 KQVYRKIGVKVEQAGKQAWDHRSQSTSTAEKLTLLLTASGQAVTSMKTYPPRPAHLWQRP 295
DB 242 LRIYRAIGTRIQGGPEAYRQRISTSKAATIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

RESULT 14
US-08-579-667-6
Sequence 6, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grilli, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-6

Query Match 20.7%; Score 318.5; DB 1; Length 410;
Best Local Similarity 28.6%; Pred. No. 1.3e-26;

Matches 82; Conservative 57; Mismatches 129; Indels 19; Gaps 5;
y 5 SKSPATASTLFDKATRRSVLMYAWCRHDDVDVDDTLGFDADQSSOMPEORLOOLEWK 64
b 130 AKTFYLGTMLTERRRAIWAIVWCRRTDDELVDGN-----ASHITPQALDRWEDR 181
y 65 TRQAYAGSQMHEPFAFAFOEVAMAH--DIAPAYAPDHLGFGFAMDVRETRYLTLDOTLRYC 122
b 182 LEDVFGS-RPFDMLDAALSDTSVQFPVDIQPFR--DWIEGMDLKSRYRNFDELYLYC 238
y 123 YHVAGVGLMAQIMGVDRNA-----TLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
b 182 LEDVFGS-RPFDMLDAALSDTSVQFPVDIQPFR--DWIEGMDLKSRYRNFDELYLYC 238
y 123 YHVAGVGLMAQIMGVDRNA-----TLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
b 239 YYVAGTVGLMSVPIMGIAPIAPDSKATTENVVNAALALGIANQLNILDVGEDARRGRVILP 298
y 177 ESMLEEGITKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAOLPLRSAAWAIATAK 236
b 239 YYVAGTVGLMSVPIMGIAPIAPDSKATTENVVNAALALGIANQLNILDVGEDARRGRVILP 298
y 177 ESMLEEGITKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAOLPLRSAAWAIATAK 236
b 299 QDELAHAGLSDDDI FAGKVTDKWRSPMKQIQARKEPDEAEEGVTQLSSASRWPVWASL 358
y 237 QVYRKIGVKVEQAGKQAWDHROSTSTAEXLTLLLTASGOAVTSRMT 283
b 299 QDELAHAGLSDDDI FAGKVTDKWRSPMKQIQARKEPDEAEEGVTQLSSASRWPVWASL 358
y 237 QVYRKIGVKVEQAGKQAWDHROSTSTAEXLTLLLTASGOAVTSRMT 283
b 359 LLVRLQILDEIANDYNNFTKRAYVSKSKLISLPANAKSLVPPTRT 405

RESULT 15
S-08-579-667-2
Sequence 2, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPES: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-579-667-2

Query Match 20.8%; Score 315.5; DB 1; Length 410;
Best Local Similarity 28.6%; Pred. No. 2.9e-26;
Matches 82; Conservative 57; Mismatches 129; Indels 19; Gaps 5;
y 5 SKSPATASTLFDKATRRSVLMYAWCRHDDVDVDDTLGFDADQSSOMPEORLOOLEWK 64
b 130 AKTFYLGTMLTERRRAIWAIVWCRRTDDELVDGN-----ASHITPQALDRWEDR 181

Db 130 AKTFYLGTMLTERRRAIWAIVWCRRTDDELVDGN-----ASHITPQALDRWEDL 181
Qy 65 TRQAYAGSQMHEPFAFAFOEVAMAH--DIAPAYAPDHLGFGFAMDVRETRYLTLDOTLRYC 122
Db 182 LEDVFGS-RPFDMLDAALSDTSVSKFPVDIQPFR--DWIEGMDLKSRYRNFDELYLYC 238
Qy 123 YHVAGVGLMAQIMGVDRNA-----TLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
Db 239 YYVAGTVGLMSVPIMGIAPIAPDSKATTENVVNAALALGIANQLNILDVGEDARRGRVILP 298
Qy 177 ESMLEEGITKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAOLPLRSAAWAIATAK 236
Db 299 QDELAHAGLSDDDI FAGKVTDKWRSPMKQIQARKEPDEAEEGVTQLSSASRWPVWASL 358
Qy 237 QVYRKIGVKVEQAGKQAWDHROSTSTAEXLTLLLTASGOAVTSRMT 283
Db 359 LLVRLQILDEIANDYNNFTKRAYVSKSKLISLPANAKSLVPPTRT 405

Search completed: February 29, 2004, 14:55:04
Job time : 10.9949 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:26:38 ; Search time 34.9969 Seconds
(without alignments)
2389.754 Million cell updates/sec

tle: US-09-941-947A-34

fect score: 1535
quence: 1 NAVGSKSFATSTLFDKTR.....VTSRMKTYPPRPRLHLMORPI 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

al number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A Geneseq_29Jan04.*

- 1: geneseq1980s.*
- 2: geneseq1990s.*
- 3: geneseq2000s.*
- 4: geneseq2001s.*
- 5: geneseq2002s.*
- 6: geneseq2003as.*
- 7: geneseq2003bs.*
- 8: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1535	100.0	296	5	Aae22315 Pantoea s
2	1535	100.0	296	6	Aao16022 Pantoea s
3	1535	100.0	296	6	Abp96689 Pantoea s
4	1368	89.1	296	2	Aaw87887 Protein e
5	1368	89.1	296	2	Aaw99097 Erwinia u
6	1368	89.1	309	2	Aaw93845 E. uredo
7	1363	88.8	296	2	Aar07467 Polypepti
8	1363	88.8	309	2	Aaw82256 C. utilis
9	1028	67.0	309	2	Aar16492 Erwinia h
10	997.5	65.0	308	2	Aar13983 Phytoene
11	997.5	65.0	308	2	Aaw01121 Phytoene
12	997.5	65.0	308	2	Aaw00172 E. herbic
13	997.5	65.0	308	2	Aaw00343 Phytoene
14	997.5	65.0	308	2	Aaw32471 Erwinia h
15	904	58.9	309	6	Abm70124 Photorhab
16	686.5	44.7	303	2	Aaw06516 Flavobact
17	686.5	44.7	303	2	Aaw69531 Flavobact
18	336.5	21.9	440	5	Aam51841 Nicotiana
19	318.5	20.7	410	2	Aaw41059 Phytoene
20	315.5	20.6	410	2	Aaw41057 Phytoene
21	311.5	20.3	410	2	Aaw41060 Phytoene
22	311.5	20.3	413	2	Aaw41058 Phytoene
23	310.5	20.2	410	5	Aam51842 Nicotiana
24	309.5	20.2	400	3	Aay44217 Soybean p
25	307	20.0	363	3	Aag10660 Arabidops

ALIGNMENTS

RESULT 1
AAE22315
ID AAE22315 standard; protein; 296 AA.

XX AC AAE22315;

XX AC (first entry)

DE Pantoea stewartii phytoene synthase (CrtB) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;
KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KW aquaculture; enzyme; phytoene synthase; CrtB.

XX Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229858P.

XX 01-SEP-2000; 2000US-0229907P.

XX (DUFO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odom JM, Picataggio SK, Rouviere PB;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35513.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by
using microorganisms having a nucleic acid molecule encoding enzymes in
the carotenoid biosynthetic pathway and which metabolize single carbon
substrates.

XX Claim 17; Page 144-145; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The
method comprises a transformed metabolising host cell, comprising
suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
encoding an enzyme in the carotenoid biosynthetic pathway, under the
control of regulatory sequences, and contacting the host cell with carbon
substrate to produce a carotenoid compound. The method is useful for
producing carotenoid compounds such as antheraxanthin and astaxanthin, by
using microorganism having a nucleic acid molecule encoding enzymes in

the carotenoid biosynthetic pathway and which metabolise single carbon substrates. The carotenoids have potent anti-oxidant properties useful in diet, and aquaculture elements. The carotenoids are also useful as intermediates in the synthesis of steroids, flavours and fragrances and compounds for potential electro-optic applications. The present sequence is Pantoea stewartii phytoene synthase (CrtB) enzyme used in the invention

Query Match 100.0%; Score 1535; DB 5; Length 296;
Best Local Similarity 100.0%; Pred. NO. 1.7e-167;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVSKSFATASTLFDKTRRSVLMYAMCRHCDVDDQTLGFHADQSSQMEQRLQQ 60
1 MAVSKSFATASTLFDKTRRSVLMYAMCRHCDVDDQTLGFHADQSSQMEQRLQQ 60
61 LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
61 LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
241 KIGYKVEQAGKQAWDHQSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
241 KIGYKVEQAGKQAWDHQSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296

RESULT 2
ID AAO16022 standard; protein; 296 AA.
AC AAO16022;
XT 20-FEB-2003 (first entry)
DE Pantoea stewartii phytoene synthase.
KW Carotenoid; crt.
OS Pantoea stewartii.
PN WO200279395-A2.
PD 10-OCT-2002.
PF 25-JAN-2002; 2002WO-US002124.
PR 26-JAN-2001; 2001US-0264329P.
PR 04-MAY-2001; 2001US-0288984P.
PR (CRG) CARGILL INC.
PI De Souza ML, Kollmann SR, May CA, Schroeder WA;
DR WPI; 2003-075455/07.
DR N-PSDB; ABL14194.

Novel isolated nucleic acid useful e.g. to engineer host cells with the ability to produce particular carotenoids and polypeptides useful in cell-free systems to make particular carotenoids.

Claim 32; Page 64-65; 74pp; English.

The invention comprises the amino acid and coding sequence of a number of carotenoid (crt)-related proteins. The crt-related DNA and protein

sequences of the invention are useful for engineering cells which are able to produce carotenoids. The present amino acid sequence represents a crt-related protein of the invention

Query Match 100.0%; Score 1535; DB 6; Length 296;
Best Local Similarity 100.0%; Pred. NO. 1.7e-167;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVSKSFATASTLFDKTRRSVLMYAMCRHCDVDDQTLGFHADQSSQMEQRLQQ 60
1 MAVSKSFATASTLFDKTRRSVLMYAMCRHCDVDDQTLGFHADQSSQMEQRLQQ 60
61 LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
61 LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
241 KIGYKVEQAGKQAWDHQSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
241 KIGYKVEQAGKQAWDHQSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296

RESULT 3
ID ABP96689 standard; protein; 296 AA.
AC ABP96689;
DT 03-JUN-2003 (first entry)
DE Pantoea stewartii phytoene synthase SEQ ID NO:10.
KW Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY; crtI; crtB; crtZ; phytoene synthase; enzyme; phytoene; carotenoid.
OS Pantoea stewartii.
PN WO2003016503-A2.
PD 27-FEB-2003.
PF 15-AUG-2002; 2002WO-US026647.
PR 15-AUG-2001; 2001US-0312646P.
PR (DUPO) DU PONT DE NEMOURS & CO E I.
PI Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;
DR WPI; 2003-268323/26.
DR N-PSDB; ACC44763.

Novel nucleic acid molecule isolated from Pantoea stewartii encoding a carotenoid biosynthetic enzyme, useful for regulating carotenoid biosynthesis in an organism.

Claim 4; Page 66; 68pp; English.

The present invention describes Pantoea stewartii carotenoid biosynthetic enzymes (I). More specifically described are the geranylgeranyl pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX), lycopene cyclase (crtI), phytoene desaturase (crtI), phytoene synthase (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating

carotenoid biosynthesis in an organism, by over-expressing (I) in an organism, such that the carotenoid biosynthesis is altered in the organism. (I) and the genes encoding (I) are useful for converting phytoene to the carotenoids, for creating recombinant organisms that have the ability to produce various carotenoid compounds, and also for enhancing or manipulating carotenoid compounds. (I) can also be used for producing gene products having enhanced or altered activity

Sequence 296 AA;

Query Match 100.0%; Score 1535; DB 6; Length 296;
 Best Local Similarity 100.0%; Pred. No. 1.7e-167; Length 296;
 Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVGSKSPATASTLFDKTRRSVLMYAWCRHCDVDDQTLGFGHADQPSQMPERLQQ 60
 1 MAVGSKSPATASTLFDKTRRSVLMYAWCRHCDVDDQTLGFGHADQPSQMPERLQQ 60

61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120
 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

181 EEEGLTKANYAAPENRQALSRAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240
 181 EEEGLTKANYAAPENRQALSRAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240

241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 4

AAW87887
 AAW87887 standard; protein; 296 AA.

AAW87887;

17-OCT-2003 (revised)
 10-MAY-1999 (first entry)

Protein encoded by the carotenoid biosynthesis gene crtB.

Carotenoid biosynthesis; astaxanthin diglucoside; crtB gene; crtB gene;
 crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;
 carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
 food additive.

Pantoea ananatis.

JP10327865-A.

15-DEC-1998.

29-MAY-1997; 97JP-00140460.

29-MAY-1997; 97JP-00140460.

(KIRI) KIRIN BREWERY KK.

(KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

WPI; 1999-099030/09.

N-PSDB; AAV84079.

New carotenoid glucoside(s) - used as food additives.

Disclosure; Page 15-16; 26pp; Japanese.

The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and

adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glycoside, in which all, or part of, carotenoid biosynthesis genes crtB, crtB, crtI, crtY, crtZ, crtX or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 296 AA;

Query Match 89.1%; Score 1368; DB 2; Length 296;

Best Local Similarity 88.9%; Pred. No. 2.8e-148;

Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 1 MAVGSKSPATASTLFDKTRRSVLMYAWCRHCDVDDQTLGFGHADQPSQMPERLQQ 60

Db 1 MAVGSKSPATASTLFDKTRRSVLMYAWCRHCDVDDQTLGFGHADQPSQMPERLQQ 60

QY 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

Db 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

QY 121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

Db 121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

QY 181 EEEGLTKANYAAPENRQALSRAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240

Db 181 EEEGLTKANYAAPENRQALSRAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240

QY 241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

Db 241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 5

AAW99097

ID AAW99097 standard; protein; 296 AA.

XX AAW99097;

17-OCT-2003 (revised)

14-MAY-1999 (first entry)

Erwinia uredovora crtB protein sequence.

Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;
 metabolite.

Pantoea ananatis.

JP11046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

07-AUG-1997; 97JP-00213648.

(KIRI) KIRIN BREWERY KK.

WPI; 1999-208113/18.

N-PSDB; AAX19117.

Beta-carotene hydroxylase - useful for preparation of xanthophylls and their metabolites.

Disclosure; Page 11-12; 17pp; Japanese.

The present invention describes beta-carotene hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crtB protein sequence

Q Sequence 296 AA;

Query Match 88.8%; Score 1363; DB 2; Length 296;

Best Local Similarity 88.5%; Pred. No. 1e-147;

Matches 262; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDIDDQTLGFFHADOPSSQMPQRLOQ 60
 14 MAVGSKSFATASKLFDKTRRSVLMYAWCRHCDVDIDDQTLGFFHADOPSSQMPQRLOQ 73
 61 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 120
 74 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 133
 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 134 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193
 181 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 240
 194 EHEGLNKENYAAPENRQALSRIARELVQEAEPYLSATAGLAGLPLRSAWAIAATAKQVYR 253
 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 8

AW82256

AAW82256 standard; protein; 309 AA.

AAW82256;

17-OCT-2003 (revised)

16-JUL-1999 (first entry)

C. utilis crtB protein.

HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtB;

Carotenoid.

Pichia jadinii.

JP10248575-A.

22-SEP-1998.

12-MAR-1997; 97JP-00058012.

12-MAR-1997; 97JP-00058012.

(KIRI) KIRIN BREWERY KK.

WPI; 1998-560727/48.

N-PSDB; AA73180.

Gene useful for increase in carotenoid production - and preparation of

carotenoid.

Example 2; Fig 8-10; 54pp; Japanese.

This invention describes a novel method for the preparation of
 carotenoids using genes and proteins isolated from *Candida utilis*. The
 invention specifically describes the isolation of a 3-hydroxy-3-
 methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence
 represents the *Candida utilis* crtB protein which is used in the method of
 the invention. (Updated on 17-Oct-2003 to standardise OS field)

Q Sequence 309 AA;

Query Match 88.8%; Score 1363; DB 2; Length 309;

Best Local Similarity 88.5%; Pred. No. 1.1e-147;

Matches 262; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

QY 1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDIDDQTLGFFHADOPSSQMPQRLOQ 60
 DB 14 MAVGSKSFATASKLFDKTRRSVLMYAWCRHCDVDIDDQTLGFFHADOPSSQMPQRLOQ 73
 QY 61 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 120
 DB 74 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 133
 QY 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 DB 134 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193
 QY 181 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 240
 DB 194 EHEGLNKENYAAPENRQALSRIARELVQEAEPYLSATAGLAGLPLRSAWAIAATAKQVYR 253
 QY 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 DB 254 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 309

RESULT 9

AAW16492

ID AAW16492 standard; protein; 309 AA.

AC AAW16492;

XX XX

17-OCT-2003 (revised)

22-JUL-1997 (first entry)

XX XX

Brwinia herbicola phytoene synthase.

Transgenic plant; selectable marker; carotenoid; pigment;

phytoene synthase; lycopene; plasmid pET0203.

XX OS

Pantoea agglomerans.

XX XX

WO9714807-A1.

XX XX

24-APR-1997.

XX XX

29-MAR-1996; 96WO-US004313.

XX XX

16-OCT-1995; 95US-00543608.

XX XX

(SEMI-) SEMINIS VEGETABLES.

XX XX

Trulson AJ, Braun CU;

XX XX

WPI; 1997-245122/22.

XX XX

N-PSDB; AAT66534.

Visual identification of transgenic plant material - from production of
 carotenoid pigment encoded by cassette containing *Erwinia* phytoene
 synthase gene, useful for selecting material for regeneration.

Example; Page 33-34; 62pp; English.

Erwinia herbicola phytoene synthase (AAW16492) catalyses the biosynthesis
 of phytoene, a precursor of the red carotenoid lycopene, from
 geranylgeranyl pyrophosphate. Its expression in transgenic plant material
 can be utilised as a means of visually distinguishing such material from
 non-transgenic cells and tissues. Novel binary vector pET0203 (ATCC
 97282) includes an expression cassette comprising the tomato E8 promoter
 (see also AAT66533), a plastid targeting signal (AAT66535) fused to the
 phytoene synthase gene (AAT66534), and a 3' non-translated region
 (AAT66536). Transgenic plant cells and tissues are identified by the
 appearance of orange colour. (Updated on 17-Oct-2003 to standardise OS
 field)

XX Sequence 309 AA;

Query Match 67.0%; Score 1028; DB 2; Length 309;
Best Local Similarity 64.9%; Pred. No. 4e-109;
Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

2Y 1 MAVGSKFATASTILFDKATRSVLMLYAWCHCDDVDVDDQTLGFHADQPSQMPERLQO 60
2b 14 MANGSKFATAAKLFDPATRSVLMLYAWCHCDDVDVDDQTLGFHADQPSQMPERLQO 73

2Y 61 LEMKTRQAYAGSQMHBPAPAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 120
2b 74 LRTLTAAFEAGAQMDPAPAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 133

2Y 121 YCYHVAGVGLMAQINGVNDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
2b 134 YCYHVAGVGLMAQINGVNDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193

2Y 181 BEEGLTKANYAAPENRQALSRIGRLVREAEPPYVSSMAGLQPLRSAWAIATAKQVYR 240
2b 194 QDAGLTPENYAARENRPALARWR-RUJIDAAEPYVSSQAGLHDLPRCAWAIATARSVYR 253

2Y 241 KIGVVCAGQAWDHRQSTSTAKLTLLLTASQAVTSRMKTYPRPAHLWORPI 296
2b 254 EIGIKVKAAGSAMDREQHTSKGEKIAMLAAPQVIRAKTRTTPRPAGLWQRPV 309

RESULT 10
AAW01121 standard; protein; 308 AA.

XX AC AAW01121;
XX DT 16-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 11-DEC-1996 (first entry)
XX DE Phytoene synthase.
XX KW GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
XX KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
XX KW pigment; food colourant; chloroplast transit peptide; increase yield;
XX KW tobacco ribulose bis-phosphate carboxylase-oxygenase.
XX OS Pantoea agglomerans.
XX PN US5530188-A.
XX PD 25-JUN-1996.
XX PF 21-JUL-1993; 93US-00095726.
XX PR 02-MAR-1990; 90US-00487613.
XX PR 18-MAY-1990; 90US-00525551.
XX PR 03-AUG-1990; 90US-00562674.
XX PR 28-FEB-1991; 91US-00662921.
XX PR 30-OCT-1991; 91US-00785566.
XX PA (STAD) AMOCO CORP.
XX PI Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
XX PI Ausich RL;
XX DR WPI; 1996-308823/31.
XX DR N-PSDB; AAT40791.

Query Match 65.0%; Score 997.5; DB 2; Length 308;
Best Local Similarity 63.9%; Pred. No. 1.3e-105;
Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

QY 1 MAVGSKFATASTILFDKATRSVLMLYAWCHCDDVDVDDQTLGFHADQPSQMPERLQO 60
Db 14 MANGSKFATAAKLFDPATRSVLMLYAWCHCDDVDVDDQTLGFHADQPSQMPERLQO 73

QY 61 LEMKTRQAYAGSQMHBPAPAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 120
Db 74 LRTLTAAFEAGAQMDPAPAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 133

QY 121 YCYHVAGVGLMAQINGVNDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
Db 134 YCYHVAGVGLMAQINGVNDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193

QY 181 BEEGLTKANYAAPENRQALSRIGRLVREAEPPYVSSMAGLQPLRSAWAIATAKQVYR 240
Db 194 QDAGLTPENYAARENRPALARWR-RUJIDAAEPYVSSQAGLHDLRRSAWAIATARSVYR 252

QY 241 KIGVVCAGQAWDHRQSTSTAKLTLLLTASQAVTSRMKTYPRPAHLWORPI 296
Db 253 EIGIKVKAAGSAMDREQHTSKGEKIAMLAAPQVIRAKTRTTPRPAGLWQRPV 308

RESULT 11
AAW01121 standard; protein; 308 AA.

XX AC AAW01121;
XX DT 16-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 11-DEC-1996 (first entry)
XX DE Phytoene synthase.
XX KW GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
XX KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
XX KW pigment; food colourant; chloroplast transit peptide; increase yield;
XX KW tobacco ribulose bis-phosphate carboxylase-oxygenase.
XX OS Pantoea agglomerans.
XX PN US5530188-A.
XX PD 25-JUN-1996.
XX PF 21-JUL-1993; 93US-00095726.
XX PR 02-MAR-1990; 90US-00487613.
XX PR 18-MAY-1990; 90US-00525551.
XX PR 03-AUG-1990; 90US-00562674.
XX PR 28-FEB-1991; 91US-00662921.
XX PR 30-OCT-1991; 91US-00785566.
XX PA (STAD) AMOCO CORP.
XX PI Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
XX PI Ausich RL;
XX DR WPI; 1996-308823/31.
XX DR N-PSDB; AAT40791.

Increasing prodn. of total carotenoid(s) in a higher plant - by transforming with vector encoding chloroplast transit peptide operably linked to the Erwinia herbicola lycopene cyclase structural gene.

Example 4; Fig 4; 99pp; English.

The present sequence is that of the Erwinia herbicola phytoene synthase which acts on geranylgeranyl pyrophosphate (GGPP) (produced by the action of GGPP synthase (AAW0119) on isopentenyl pyrophosphate and dimethylallyl pyrophosphate). GGPP is a 20-carbon atom precursor of phytoene, the first carotenoid in the carotenoid biosynthesis pathway. Production of total carotenoids in a plant can be increased by transforming the plant with DNA encoding enzymes involved in the biosynthesis pathway, in particular the lycopene cyclase gene (AA40795). Lycopene is prepared biosynthetically from phytoene through four sequential dehydrogenation reactions which can be carried out by a single dehydrogenase (AA40793) in Erwinia sp. Beta-carotene is produced by the action of lycopene cyclase on lycopene. A chloroplast transit peptide of the tobacco ribulose bis-phosphate carboxylase-oxygenase gene (see AA40794) is operatively linked in frame to the 5' end of the lycopene cyclase structural gene. This leads to increased production of total carotenoids in the chloroplast of transformed plants as compared to native, non-transformed plants of the same type. Beta-carotene is an effective and apparently harmless food colourant and is also in the pathway for biological synthesis of further C40 carotenoids such as zeaxanthin and zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 308 AA;

Query Match 65.0%; Score 997.5; DB 2; Length 308;
Best Local Similarity 63.9%; Pred. No. 1.3e-105;
Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;
1 MAVGSKSPATASTLFDKTRRSVLMYAWCHCDVDVDDQTLGFHADQPSQMPQBLQQ 60
14 MANGSKSPATAAKLFDPATRSVLMYTWCHRCDDVDDQTHGFASAAAEEATQRLAR 73
61 LEMKTRQYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYLTLDPTLR 120
74 LRTLTAAFEAGMOPDPAFAQEVVALTHGTPRMALDHLDFGFMADVQTRYVTFEPTLR 133
121 YCYHVAGVVGMLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPRESWL 180
134 YCYHVAGVVGMLMARVGMVGRDRLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPRESWL 193
181 EEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQPLPSAWAIATAKOVYR 240
253 EIGIKVKAAGGSAMDWRQHTSKGKIAMLAAPQGVIRAKTTRVTPRPAGLWQRPV 308

SULT 12
AW00172

AAW00172 standard; protein; 308 AA.

AAW00172;
16-OCT-2003 (revised)
25-MAR-2003 (revised)
17-OCT-1996 (first entry)

E. herbicola phytoene synthase encoded by parC285.

E. herbicola; geranylgeranyl pyrophosphate synthase; parC376;
GGPP synthase; biosynthesis; carotenoid; lycopene;
farnesyl pyrophosphate; phytoene; fpp; isopentenyl pyrophosphate; IPP;
tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;
food colourant; herbicide; norflurazon.

Pantoea agglomerans.

US5530189-A.

25-JUN-1996.

22-JUL-1993; 93US-00096043.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

30-OCT-1991; 91US-00785568.

(STAD) AMOCO CORP.

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
Ausich RL;

WPI; 1996-308824/31.

N-PSDB; AAT37093.

DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for prodn. of lycopene, and to produce transgenic plants resistant to norflurazon.

Example 4; Fig 4; 87pp; English.

This sequence represents Erwinia herbicola phytoene synthase encoded by plasmid pARC285. Phytoene synthase is an enzyme which is involved in the biosynthesis of carotenoids, esp. lycopene, from the ubiquitous precursor, farnesyl pyrophosphate. In E. herbicola, phytoene has been found to be formed biosynthetically in a two-step process. The initial step is the condensation of farnesyl pyrophosphate (fpp) and isopentenyl pyrophosphate (IPP) to form GGPP. This reaction is catalysed by GGPP synthase. This first step is immediately followed by a tail to tail dimerisation of GGPP, catalysed by the enzyme phytoene synthase, to form phytoene. Lycopene is produced from phytoene by the catalytic action of phytoene dehydrogenase-4H. The genes encoding components of the lycopene biosynthesis pathway may be used to transform a host cell for the commercial production of lycopene which is used as a food colourant. Plants transformed with the phytoene dehydrogenase-4H coding sequence are protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 308 AA;

Query Match 65.0%; Score 997.5; DB 2; Length 308;
Best Local Similarity 63.9%; Pred. No. 1.3e-105;
Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

1 MAVGSKSPATASTLFDKTRRSVLMYAWCHCDVDVDDQTLGFHADQPSQMPQBLQQ 60

14 MANGSKSPATAAKLFDPATRSVLMYTWCHRCDDVDDQTHGFASAAAEEATQRLAR 73

61 LEMKTRQYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLEGFAMDVRETRYLTLDPTLR 120

74 LRTLTAAFEAGMOPDPAFAQEVVALTHGTPRMALDHLDFGFMADVQTRYVTFEPTLR 133

121 YCYHVAGVVGMLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPRESWL 180

134 YCYHVAGVVGMLMARVGMVGRDRLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPRESWL 193

181 EEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQPLPSAWAIATAKOVYR 240

253 EIGIKVKAAGGSAMDWRQHTSKGKIAMLAAPQGVIRAKTTRVTPRPAGLWQRPV 308

RESULT 13

AAW00343	D	AAW00343 standard; protein; 308 AA.	
AC	C	AAW00343;	
XT	T	16-OCT-2003 (revised)	
TT	T	25-MAR-2003 (revised)	
XT	T	10-DEC-1996 (first entry)	
EX	E	Phytoene synthase.	
XX	W	Wild type; N-terminally truncated; geranylgeranyl pyrophosphate; GGPP;	
XX	W	GGPP synthase; phytoene synthase; carotenoid; phytoene; pARC306A;	
XX	W	pARC489D; transit peptide; tobacco; chloroplast; skin disorder; ss.	
XX	X	Pantoea agglomerans.	
XX	X	US5545816-A.	
XX	D	13-AUG-1996.	
XX	F	19-JUL-1993; 93US-00093577.	
XX	X	02-MAR-1990; 90US-00487613.	
XX	R	18-MAY-1990; 90US-00525551.	
XX	R	03-AUG-1990; 90US-00562674.	
XX	R	28-FEB-1991; 91US-00662921.	
XX	R	30-OCT-1991; 91US-00785569.	
XX	A	(STAD) AMOCO CORP.	
XX	X	Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;	
XX	PI	Ausich RL;	
XX	R	WPI; 1996-383742/38.	
XX	R	N-PSDB; AAT41743.	
XX	XT	Transformed plants containing DNA encoding Erwinia herbicola enzymes -	
XX	PT	esp. geranylgeranyl pyrophosphate synthase and phytoene synthase, allows	
XX	PT	large scale production of phytoene.	
XX	XX	Disclosure; Fig 4; 61pp; English.	
XX	CC	This sequence encodes wild type phytoene synthase. Phytoene synthase is	
XX	CC	an enzyme which, along with geranylgeranyl pyrophosphate (GGPP) synthase	
XX	CC	(see also AA41741-42), catalyze the formation of carotenoids,	
XX	CC	particularly phytoene. An N-terminally truncated form of GGPP synthase	
XX	CC	which has the first thirteen amino acids replaced by four heterologous	
XX	CC	amino acids derived from the plasmid pARC306A was found to be about twice	
XX	CC	as active as the wild type enzyme. A C-terminal truncated protein was	
XX	CC	found to have even higher activity and was cloned into plasmid pARC489D.	
XX	CC	The phytoene synthase coding sequences may be attached to the transit	
XX	CC	peptide coding sequence given in AA41744, for transport into tobacco	
XX	CC	chloroplasts. This allows the tobacco plants to produce high levels of	
XX	CC	phytoene for use in the treatment of skin disorders. (Updated on 25-MAR-	
XX	CC	2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS	
XX	CC	field)	
XX	XX	Sequence 308 AA;	
XX	XX	Query Match 65.0%; Score 997.5; DB 2; Length 308;	
XX	XX	Best Local Similarity 63.9%; Pred. No. 1.3e-105;	
XX	XX	Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;	
XX	QY	1 MAVGSKSFATATLFDKATRSVLMYAKCHCDVDVDDOTLGHADOPSSQMPQRLQQ 60	
XX	DB	14 MANGSKSFATAKLFDPATRSVLMYAKCHCDVDVDDOTLGHADOPSSQMPQRLQQ 73	
XX	QY	61 LEMKTRQVAGSQMEPFAFAFQVEMAHDIAPAYAFDHLGFMADVRETRYTLTDTLR 120	
XX	DB	74 LRTLLAFAFEAGEMQDPFAFAFQVEMAHDIAPAYAFDHLGFMADVRETRYTLTDTLR 133	
XX	QY	121 YCYHVAGVGLMMAQINGVRDNTLDRACDLGLAFQLTNNARDIIDDAIDRCYLPAML 180	
DB	Db	134 YCYHVAGVGLMMAQINGVRDNTLDRACDLGLAFQLTNNARDIIDDAIDRCYLPAML 193	
QY	QY	181 EEBGLTKANYAAPENQALSRAGRLVRRAEPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240	
DB	Db	194 ODAGLAPENTAARENRPALAWR-RLIDAAEPYIISQAGLHDLRRSAWAIATASVYR 252	
QY	QY	241 KIGVKVEQAGQAWHDHQSTSTAEKUTLLLTASQAVTSMKTYPPRPAHLWOPPI 296	
DB	Db	253 EIGIKVKAAGSAWDRRQHTSKGEKIAMLMAPQVIRAKTTTVPTRPAGLWOPV 308	
RESULT 14			
AAW32471			
ID	AAW32471	standard; protein; 308 AA.	
XX	AC	AAW32471;	
XX	XX	17-OCT-2003 (revised)	
DT	DT	25-MAR-2003 (revised)	
DT	DT	15-JAN-1998 (first entry)	
XX	XX	Erwinia herbicola phytoene synthase.	
XX	XX	Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;	
XX	XX	lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;	
XX	XX	yeast; plant; vitamin A; cancer.	
XX	XX	Pantoea agglomerans.	
XX	XX	US5656472-A.	
XX	PD	12-AUG-1997.	
XX	XX	07-JUN-1995; 95US-00473512.	
XX	XX	02-MAR-1990; 90US-00487613.	
XX	PR	18-MAY-1990; 90US-00525551.	
XX	PR	03-AUG-1990; 90US-00562674.	
XX	PR	28-FEB-1991; 91US-00662921.	
XX	PR	21-JUL-1993; 93US-00095726.	
XX	XX	(STAD) AMOCO CORP.	
XX	XX	Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;	
XX	PI	Brinkhaus FL;	
XX	DR	WPI; 1997-414592/38.	
XX	DR	N-PSDB; AAT91543.	
XX	XX	DNA encoding Erwinia herbicola lycopene cyclase - for producing	
XX	PT	recombinant enzyme, and transgenic organisms with increased beta-carotene	
XX	PT	levels.	
XX	XX	Example 8; Fig 4; 102pp; English.	
XX	XX	A novel DNA molecule has been isolated which encodes an Erwinia herbicola	
XX	CC	lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA	
XX	CC	molecule comprises at least 1125 bp and is present in the plasmids	
XX	CC	pARC147, pARC1509, pARC1510 and pARC1520. The present sequence represents	
XX	CC	amino acid corresponding to the structural gene for phytoene synthase.	
XX	CC	The new DNA molecule can be used to produce the recombinant enzyme and	
XX	CC	transgenic organisms, e.g. Yeasts or plants, with increased beta-carotene	
XX	CC	levels. Beta-carotene is used as a colourant in margarine and butter and	
XX	CC	as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-	
XX	CC	MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to standardise OS	
XX	CC	field)	
XX	XX	Sequence 308 AA;	
XX	XX	Query Match 65.0%; Score 997.5; DB 2; Length 308;	
XX	XX	Best Local Similarity 63.9%; Pred. No. 1.3e-105;	
XX	XX	Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;	

1 MAVGSKSFATSTLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRLQQ 60
14 YAVGSKSFATSKLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRLQQ 73
61 LEMKTRQYAGSQMHEPAPAFQEVAMAHDTAPAYADHLEGPAMDVRETRYLTLDOTLR 120
74 LRTLLAAFEAGMOPDPAFAQFQVAVTHGTERMALDHLGDFAMDVAVQRYVTFEOTLR 133
121 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
134 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193
181 EEEGLTKANYAAPENRQALSRAGLVREAPYVYSSMAGLAQLPLRSAAWAIATAKQVYR 240
194 ODAGLAPENYAARENRPALARWR-REIDAEFYVYSSOAGLHDLRRRSAAWAIATAKQVYR 252
241 KIGVKVEAGKQAMDRHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
253 EIGIKVKAAGGSAMDRHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 309

RESULT 15

M70124

ABM70124 standard; protein; 309 AA.

ABM70124;

20-NOV-2003 (first entry)

Photorhabdus luminescens protein sequence #3221.

Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
detection; food; gene expression; plant; animal; microorganism; toxin;
antibiotic; biopesticide; virulence factor; disease model; plague;
whooping cough.

Photorhabdus luminescens.

WC200294867-A2.

28-NOV-2002.

07-FEB-2002; 2002WC-IB003040.

07-FEB-2001; 2001FR-00001659.

(INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.

Duchaud E, Taouit S, Glaser P, Frangeul L, Kunst P, Danchin A;
Buchrieser C;

WPI; 2003-148459/14.

Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
useful e.g. as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 3221; 1205pp; French.

The invention relates to the isolation of genes and their encoded
proteins from Photorhabdus luminescens. The isolated sequences are
sources of probes and primers for detecting the genome of P. luminescens
and related species; to study polymorphisms; for gene analysis and for
detection/amplification of the genes. Antibodies (Ab) raised against the
polypeptides encoded by the genes are used for detection/identification
of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
carry a gene-containing vector are used to select compounds that
modulate, regulate, induce or inhibit expression of the genes in plants,
animals or microorganisms other than P. luminescens and are able to alter
response or sensitivity to toxins and antibiotics produced by P.
luminescens. Cells transformed to express the genes are useful for
recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX

SQ Sequence 309 AA;

Query Match 58.9%; Score 904; DB 6; Length 309;

Best Local Similarity 58.6%; Pred. No. 7.6e-95;

Matches 173; Conservative 42; Mismatches 80; Indels 0; Gaps 0;

QY 1 MAVGSKSFATSTLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRLQQ 60

DB 13 MEQSKSFASVTRLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRLQQ 72

QY 61 LEMKTRQYAGSQMHEPAPAFQEVAMAHDTAPAYADHLEGPAMDVRETRYLTLDOTLR 120

DB 73 LOYLTQAYDGLMTEPAPAFQEVAMAHDTAPAYADHLEGPAMDVRETRYLTLDOTLR 132

QY 121 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

DB 133 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 192

QY 181 EEEGLTKANYAAPENRQALSRAGLVREAPYVYSSMAGLAQLPLRSAAWAIATAKQVYR 240

DB 193 HQEGLMPTDLYTENRPAALRVASRLIVEASVYTSALTGLVGLPLRSAAWAIATAKQVYR 252

QY 241 KIGVKVEAGKQAMDRHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295

DB 253 EIGIKVKAAGGSAMDRHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 307

Search completed: February 29, 2004, 14:44:06

Job time : 39.0969 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on on: February 29, 2004, 14:33:49 ; Search time 40.4556 Seconds
(without alignments)
3837.172 Million cell updates/sec

itle: US-09-941-947A-32
fect score: 2606
quence: 1 MKPTTVIGAGFGLAALRL.....IPGVIGSAKATAGLMLDLI 492

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

aximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

itabase : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_plant:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_virtebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2606	100.0	492	2 Q8GCSI	Q8GCSI pantoea ste
2	2453	94.1	492	2 Q47845	Q47845 pantoea agg
3	2298	88.2	492	2 Q8VUJ5	Q8VUJ5 pantoea agg
4	1749.5	67.1	506	2 Q9KIX2	Q9KIX2 bradyrhizob
5	1713	65.7	504	2 Q93C17	Q93C17 xanthobacte
6	1627	62.4	501	2 Q9LH4	Q9LH4 paracoccus
7	1595	61.2	494	2 P94790	P94790 flavobacter
8	1415	54.3	526	2 Q06757	Q06757 erythrobact
9	1099.5	42.2	511	2 Q9JP98	Q9JP98 rhodocyclu
10	1093.5	42.0	513	2 Q840T3	Q840T3 rhodocyclu
11	1051	40.3	548	16 Q9RW08	Q9RW08 deिनococcus
12	984.5	37.8	514	2 Q8KZ28	Q8KZ28 uncultured
13	752	28.9	517	2 Q8ZGE4	Q8ZGE4 heliobacill
14	706.5	27.1	511	16 Q8RNM0	Q8RNM0 oceanobacil
15	693.5	26.6	537	16 Q7UF97	Q7UF97 rhodopirell
16	688.5	26.4	498	16 Q88SW3	Q88SW3 lactobacill

17	638.5	24.5	507	2 Q9EXL0	Q9EXL0 streptomyce
18	636	24.4	588	2 Q83X01	Q83X01 streptomyce
19	634.5	24.3	534	16 Q7UJB6	Q7UJB6 rhodopirell
20	631.5	24.2	579	3 Q9Y7H8	Q9Y7H8 mucor circi
21	629.5	24.2	523	16 Q9R1X9	Q9R1X9 streptomyce
22	624	23.9	502	16 Q99R76	Q99R76 staphylococ
23	624	23.9	502	16 Q8NU06	Q8NU06 staphylococ
24	617	23.7	506	2 Q07855	Q07855 staphylococ
25	616.5	23.7	513	16 Q93HP2	Q93HP2 streptomyce
26	604	23.2	499	16 Q8YKN8	Q8YKN8 anabaena sp
27	592	22.7	499	2 Q44318	Q44318 anabaena sp
28	568.5	21.8	499	2 Q9K566	Q9K566 mycobacteri
29	567	21.8	490	16 Q8Y8A8	Q8Y8A8 listeria mo
30	562	21.6	582	3 Q13506	Q13506 phaffia rho
31	557	21.4	582	3 Q72858	Q72858 phaffia rho
32	549.5	21.1	514	17 Q27835	Q27835 methanobact
33	543.5	20.9	579	3 Q8X0Z0	Q8X0Z0 gibberella
34	541.5	20.8	512	17 Q9HPD8	Q9HPD8 halobacteri
35	534	20.5	544	2 Q8KP34	Q8KP34 agromyces m
36	533	20.5	494	16 Q8ENM2	Q8ENM2 oceanobacil
37	510	19.6	499	16 Q8XK44	Q8XK44 clostridium
38	489	18.8	488	16 Q92D21	Q92D21 listeria in
39	482	18.5	497	16 Q99R73	Q99R73 staphylococ
40	482	18.5	497	16 Q8NU03	Q8NU03 staphylococ
41	466	17.9	528	2 Q9KK84	Q9KK84 brevibacter
42	461	17.7	487	2 Q7WT75	Q7WT75 marine bact
43	460	17.7	488	2 Q7WT72	Q7WT72 marine bact
44	433.5	16.6	536	17 Q9HP88	Q9HP88 halobacteri
45	429	16.5	538	16 Q8FRW8	Q8FRW8 corynebacte

ALIGNMENTS

RESULT 1

ID Q8GCSI PRELIMINARY; PRT; 492 AA.
AC Q8GCSI;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phytoene desaturase.
GN CRTI.
OS Pantoea stewartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Pantoea.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8200;
RA deSouza W.L., Kollmann S.R., Schroeder W.A.;
RT "Carotenoid Biosynthesis (WO 02/079395 A2).";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY166713; AAN85599.1;
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR Pfam; PF01593; Amino oxidase; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SSEQUENCE 492 AA; 54836 MW; F3DCD224547A5FBA CRC64;

Query Match 100.0%; Score 2606; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.8e-203; Indels 0; Gaps 0;
Matches 492; Conservative 0; Mismatches 0;

QY 1 MKPTTVIGAGFGLAALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTFDAGTVDTP 60

Db 1 MKPTTVIGAGFGLAALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTFDAGTVDTP 60

QY 61 SAIELEFALAGKQLDYVELLPVTPFFYFLCHESKVFYNDQALQLENAIQCFPRDVG 120

Matches	432;	Conservative	26;	Mismatches	33;	Indels	0;	Gaps	0;
---------	------	--------------	-----	------------	-----	--------	----	------	----

1	MKPTTVIGAGFGGLAALRLQAQIPVLLLEQDKPGGRAYVYQGGFTDAGPTVITDP	60
1	MNRFTVIGAGFGGLAALRLQASGVPTLLEQDKPGGRAYVYQGGFTDAGPTVITDP	60
61	SAIEELPALAGHOLKDYVELLPVTPFFRLCWESKVFNDNDQAOLEAQIQCPNPRVAG	120
61	SAIEELFTLAGKKSDYVELMPVPPFYELCWESKVFSDNDQPALEAQIAAFNPRVVG	120
121	YRAFIDYSRAVFNEGYKLGTVPFLSFRDMLRAAPLAKLQAWRSVYSKVAGYIEDEHLR	180
121	YREFLAYSRAVFAEGYKLGTVPFLLSFRDMLRAAPLAKLQAWRSVYSKVASYIEDEHLR	180
181	QAFSFSLLVCGNPATSSIVTLHALERGWVFPFGGTGALVNGMIKLPDLGGGFWL	240
181	QAFSFSLSLVCGNPATSSIVTLHALERGWVFPFGGTGALVQGMVKLPFDLGGGYEL	240
241	NARVSHMETVGDKIQAOLEDGRGFETCAVASNADVHTVTRDLSLGHPAAXQAQKLQSK	300
241	NASVARLETOENRITAVHLKGRVPTTAVASNADVHTVTRDLSLGHPSAQGSLSQNK	300
301	RMSNSFLVLYFGLNHHHQLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP	360
301	RKSNLSLFTVYFGLNHHHQLAHTVCFGPYRELIDEIFNKDGLAEDFSLYLHAPCVTDP	360
361	SLAPEGCGSYVYLVAFVPHLGTANLDWAVEGPRLDRIFDYLEQHYMPEGILRSOLVTHRMPT	420
361	SLAPEGCGSYVYLVAFVPHLGTADLDNAVEGPRLDRIFDYLEQHYMPEGIURSQLVTHRIPT	420
421	PFDRDELNAWQGSFVSVEPILTOSAMFRPNRDKHIDNLVYVGAGTHPGAGIPGVIGSA	480
421	PFDRDELNAVQGSFVSVEPILTOSAMFRPNRDKINNLVYVGAGTHPGAGIPGVIGSA	480
481	KATAGLMLLEDL	491
481	KATAGLMLLEDL	491

SUPLY	4
KIX2	
QKIX2	PRELIMINARY; PRT; 506 AA.
01-OCT-2000	(TRMBLrel. 15, Created)
01-OCT-2000	(TRMBLrel. 15, Last sequence update)
01-JUN-2003	(TRMBLrel. 24, Last annotation update)
Phytoene desaturase.	
CXTI	
Bradyrhizobium sp. ORS278.	
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;	
Bradyrhizobiaceae; Bradyrhizobium.	
NCBI_TaxID=114615;	
[1]	
SEQUENCE FROM N.A.	
STRAIN=ORS278;	
MEDLINE=20309720;	PubMed=10851005;
Hannibal L., Lorquin J., Augles d'Ortoli N., Garcia N.,	
Chaintreuil C., Masson-Bolvin C., Dreyfus B., Giraud E.;	
"Isolation and characterization of the canthaxanthin biosynthesis	
genes from the photosynthetic bacterium Bradyrhizobium sp. strain	
ORS278."	
J Bacteriol.	182:3850-3853 (2000).
EWEL; AF218415; RA78201.1; -.	
GO; GO:0016491; F:oxidoreductase activity; IEA.	
GO; GO:0009058; P:biosynthesis; IEA.	
GO; GO:0006118; P:electron transport; IEA.	
InterPro; IPR002937; Amino oxidase.	
InterPro; IPR008150; Bac phytoene dh.	
InterPro; IPR008151; Phytin dehydro.	
Pham; PF01593; Amino oxidase; 1.	
ProDom; PD139017; Phytin dehydro; 1.	
PROSITE; PS00982; PHYTOENE DH; 1.	
SEQUENCE	506 AA; 56223 MW; 435FP9495FDEED3 CRC64;

Query Match	67.1%	Score 1749.5	DB 2	Length 506
Best Local Similarity	67.8%	Pred. No. 1.4e-133		
Matches 332	Conservative 55	Mismatches 102	Indels 1	Gaps 1
QY	2	KPTTVIAGTGGGLALATRLQAAQIPVLLLEQRDKCGRAYVYQEGFTDAGPTVITOPS	61	
DB	9	KTAVIWSGGFGGLSLATRLQASAGIATTLVEKRDKCGRAYVYQEGFTDAGPTVITOPT	68	
QY	62	AIIEELFALACKQKNDYVELLEPVTFPFYRLCWESKVFNYNDQALQEAQIQOQFNPRDVAQY	121	
DB	69	CLQELPALSGEKLNVYVELNPVSPFYQLRWEDGATEDYVNDQAEALERQIAAFCPADVDGY	128	
QY	122	RAPLDYGRAVFNBYGKLGTVPFSLPKOMLRAAPQALAKLQAWRSVYKVGAYIEDEHLRQ	181	
DB	129	RRFRSYSERLIEGYVKLGHPVDPDFRSVMRVAPQVALQSYRSYKVSQVSVSDEHLRQ	188	
QY	182	AFSPHSLLVGNGNPATSSIYTLTHALERNGYVFPFGGTGALVN-CMIKLFODLGEVVL	240	
DB	189	AFSPHSLLVGNGNPATSSIYALTHALERNGVWFPFGGTGALINKGLVOLFKDLGEVTL	248	
QY	241	NARYSHMETVGDKIQAVOLEDEGRRFETCAVSNADVVHYTRDLLSOHPAAAFQAQKLQSK	300	
DB	249	STSVGRITETANGRVSAVVAEDGRRFADIVASNADVVHYTRDLLKDEPLARPTAQSLMRK	308	
QY	301	RMSNSLFYPLGNHHEHDLAHTVCFGPYRELITHEIFNHDGLAEDPSLYLHAPCVTDP	360	
DB	309	RFSMSLFYVPLGRRRHPELKHIIILFGARYRELINIEFKGPALPDFSFLYHAPSVDTP	368	
QY	361	SLAPEGCGSYVFLAPVPHLGTANLDWAVEGPRLRDSIFDYLEQHYMGELRSOLVTHRMFT	420	
DB	369	SLAPQCGCSTYYVSPVPHLAAAPIDSVSEGFYRDRILDYLEARILFGLKSLDLCRIPT	428	
QY	421	PDFDRDELNAWQCSAFSVEPILTQSAWFRPNRDKHIDNLVLVGATHPGAGIPGVIGSA	480	
DB	429	QDQFWTELNAHLSAFSLEPILTQSAVFRAHNADDKIGLXLVGACTHPGAGIPGVIGSA	488	
QY	481	KATAGLMLEED	490	
DB	489	KATARVILEED	498	

```

RESULT 5
Q93CI7 PRELIMINARY; PERT; 504 AA.
ID Q93CI7
AC Q93CI7
DT 01-DEC-2001 {TrEMBLrel. 19, Created}
DT 01-DEC-2001 {TrEMBLrel. 19, Last sequence update}
DT 01-JUN-2003 {TrEMBLrel. 24, Last annotation update}
DE Phytocene dehydrogenase.
GN CRT1.
OS Xanthobacter sp. (strain Py2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Xanthobacter.
ON NCBI_TaxID=78245;
RX [1]
RN SEQUENCE FROM N.A.
RP
RC STRAIN=Py2;
RA Larsen R.A., Metcalf W.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF403848; AAL02000.1; -.
DR GQ; GO:0016491; P:oxidoreductase activity; IEA.
DR GQ; GO:0009058; P:biosynthesis; IEA.
DR GQ; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac_phytoene dh.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 504 AA; 55556 MW; CA848D11A607A828 CRC64;

```

[illegible]

Best Local Similarity	63.58	Pred No. 1-2e-123	Indels	Gaps
Matches 311	Conservative	58	Mismatches 121	
QY	2	KPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQSQGTFDAGPTVITDPS	61	
Db	9	KTAIVIGAGFGGLALAIRLQSGIATTLVEARDKPGGRAYVWHQDQGHVFDAGPTVITDPS	68	
QY	62	ALBELFALAGKQADXYVELLPVTFYFLRWESGKVFYNDQQAQLBAQIQQNPDRVAGY	121	
Db	69	AKLELWALTQGDWARDVTLMFVSPFTRLMPGKVFYDYNADQLERQIAQNPDDLEGY	128	
QY	122	RAFLDYGRAVFNBYGLKGTVPFLSPKMDLRAAPQLAKLQAWRSYVSKVAGYIEDEHLRQ	181	
Db	129	RRFRDYAEVYQEGYVGLGTVPFLKLGQMLKAAPALMKLEAYNSVEAKVATFIKDPYLRQ	188	
QY	182	AFSPHSLVVGNNPATSSITYLTHALERWGWYFPRGGTGALVNGMIKLFQDLGGEVVLN	241	
Db	189	AFSYHTLLVGNPFSTSSIVALLHALERGGVWFAGGNTQNLVAGHVAUFERPQGOMLIN	248	
QY	242	ARVSHMETVGDKIQAQVLEBGRFETCAVASNADVVHTYREDLLSQHPAAAKQAKKLQSKR	301	
Db	249	AKVARIETEGARTTGVTLADGRSLRADNVASNGDVVHNYRDLLGHTARGCSRAKSLDRK	308	
QY	302	MNSLFLVYFGLNHHHQLAHVTTCGPRYRELIIHFHHDGLAEDFSLYLHAPCVITDPS	361	
Db	309	WMSLFLVHLFGLRRAPKDIARHTILFGPRYKELVNEIFKGPKLAEFSLYXHSPTCTDPS	368	
QY	362	LAPGCGSYVVLAVPVPHLGTANLDWAVEGPRLDRIFDVLEOHYMPGLRSQLVTHRMFT	421	
Db	369	MAPPQMSHYVLAVPVPHLSRAEIDWAVEGPRYADRILAFLEELIPLNLRANLTRTITFT	428	
QY	422	FDPRDELNAWGSAFSVEPILITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAK	481	
Db	429	ADFASELNAHGSAFSVEPILITQSAWFRPHNRDKTIRNFYLVGAGTHPGAGIPGVIGSAK	488	
QY	482	ATAGIMLEDL	491	
Db	489	ATAQVMLSDL	498	
RESULT 7				
294790 PRELIMINARY; PRT; 494 AA.				
ID	P94790			
AC	P94790			
DT	01-MAY-1997 (Tremblrel. 03, Created)			
DT	01-MAY-1997 (Tremblrel. 03, Last sequence update)			
DE	01-JUN-2003 (Tremblrel. 24, Last annotation update)			
DE	Phycene desaturase.			
GN	CRTI.			
OS	Flavobacterium sp. ATCC 21588.			
OC	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;			
OC	Flavobacteriaceae; Flavobacterium.			
NCBI	TaxID=50286;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R1534.			
RC	MEDLINE=97186694; PubMed=9034310;			
RA	Passantes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,			
RA	van Ioon A.P.;			
RT	*Isolation and characterization of the carotenoid biosynthesis genes			
RT	of Flavobacterium sp. strain R1534.;			
RL	Gene 185-35-41[1997].			
DR	EMBL; U62808. AAC44850.1; -.			
DR	GO; GO:0016451; F:oxidoreductase activity; IEA.			
DR	GO; GO:0009058; P:biosynthesis; IEA.			
DR	GO; GO:0006118; P:electron transport; IEA.			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac.phytoene dh.			
DR	InterPro; IPR008151; Phytin dehydro.			
DR	Pfam; PF01593; Amino oxidase; 1.			
DR	ProDom; PD139017; Phytin dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE DH; 1.			
DR	SEQUENCE 494 AA; 54411 MW; CBF03BB39C1FC4D0 CRC64;			

```

RESULT 6
199RLH4
D PRELIMINARY; PRT; 501 AA.
C Q9RLH4
VT 01-MAY-2000 (TREMBLrel. 13, Created)
VT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
VT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Phytoene desaturase.
EN CRTI.
XS Paracoccus marcusii.
XC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
XC Rhodobacteraceae; Paracoccus.
XC NCBI_TaxID=59779;
IN [1]
IN SEQUENCE FROM N.A.
IC STRAIN=MH1;
IC Harker M., Hirschberg J.;
VT "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
VT MH1.";
IL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
JR EMBL; Y15112; CAB56062.1; ...
JR GO; GO:0016491; P:oxidoreductase activity; IEA.
JR GO; GO:0009058; P:biosynthesis; IEA.
JR GO; GO:0006118; P:electron transport; IEA.
JR InterPro; IPR002937; Amino oxidase.
JR InterPro; IPR008150; Bac.phytoene dh.
JR InterPro; IPR008151; Phyt. dehydro.
JR Pfam; PF01593; Amino oxidase; 1.
JR ProDom; PD139017; Phyt. dehydro; 1.
JR PROSITE; PS00982; PHYTOENE_DH; 1.
IQ SEQUENCE 501 AA; 55109 MW; F0FAED92A0E8D16 CRC64;
Query Match
62.4%; Score 1627; DB 2; Length 501;

```

```
Query Match 61.2%; Score 1595; DB 2; Length 494;
Best Local Similarity 61.9%; Pred. No. 4, 9e-121;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

Y 1 MKETTVIGAGGGLAALRILOAAGIPVLLLEBORDKPGGRAYVYQGGTFDAGPVTITDP 60
b 1 MSSAIVIGAGGGLAALRILOAAGITTTVEARDKPGGRAYVWKGDFTFDAGPVTITDP 60
Y 61 SAIEELFALAGKOLXDYVELLPVTPYRLCWESGVFNVDNDQALEAQIQOQPNRVDAG 120
b 61 DSLRELWALSGQPMERDVTLLPVSPYRLTWADGSEFYVNDDELLIQVASEFNPADVDG 120
Y 121 YRAPLDYSRAVNEGYKLGTVPFVLSFKDMLRAAPOLAKLQAWRSVYKVGAYIDEHLR 180
b 121 YRPFHDYAEVYREGYKLGTVTFPLKLGQMLNAAPALRQLAYRSVHSMVARFIQDPHLR 180
Y 181 QAFSPHLLVGNPPATSSITLTHALEREWGVPFGGTGALVNGMKLFDLGGGVVL 240
b 181 QAFSPHLLVGNPPATSSITLTHALEREWGVPFGGTGALVNGMKLFDLGGGVVL 240
Y 241 NARVSHMETVGKIQAVLEDDRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQSK 300
b 241 NARVTRIDTEGORATGVTLDDGRLQADTVASNGDVMSYRDLGHTRRGRTKAILNRQ 300
Y 301 RMSNSLFVLVEGLNHHDDLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
b 301 RMSNSLFVLVEGLNHHDDLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
Y 361 SLAPEGCGSYVVLAPVPHLGTANLDAWVEGPRLDRIIDYLEQHTMGLRSLQVTHRMFT 420
b 361 SLAPEGCMSTHYVLAPVPHLGTADWAEAPCAVARIPEELERRAIPDLRGLITVSRIFS 420
Y 421 PFDPRDELNAWGSAPSEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
b 421 PADFSTLSAHGSAFSEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
Y 481 KATAGLMLDL 491
b 481 KATAQWMLSDL 491

RESULT 8
Q6757 PRELIMINARY; PRT; 526 AA.
Q06757;
Q01-JUL-1997 (TrEMBLrel. 04, Created)
Q01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
Q01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Phytoene desaturase.
CRTI.
Erythrobaacter longus.
Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
Sphingomonadaceae; Erythrobaacter.
NCBI_TaxID=1044;
[1]
SEQUENCE FROM N.A.
MEDLINE=97311406; PubMed=9168123;
Matsumura H., Takeyama H., Hsuakabe E., Burgess J.G., Matsunaga T.;
"Cloning, sequencing and expression the carotenoid biosynthesis gene,
lycopene cyclase and phytoene desaturase, from the aerobic
phototrophic bacterium Erythrobaacter longus sp. strain Och 101 in
Escherichia coli."
Gene 189:169-174(1997).
EMBL; D83514; BAA20276.1;
GO; GO:0016491; F:oxidoreductase activity; IEA.
GO; GO:0009058; P:biosynthesis; IEA.
GO; GO:0006118; P:electron transport; IEA.
InterPro; IPR002937; Amino_oxidase.
InterPro; IPR008150; Bac phytoene dh.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino_oxidase; 1.
```

gelatinosus".
J. Biol. Chem. 269:2477-2484(1994).
[3]
SEQUENCE FROM N.A.
STRAIN=I1144;
IGarashi N., Shimada K., Matsuura K., Nagashima K.V.;
"Photosynthetic gene cluster in purple bacterium, Rubrivivax
gelatinosus"; (eds.);
(In) Garab G. (eds.);
Photosynthesis:
mechanisms and effects (Proceedings of the 11th international congress
on photosynthesis), pp.4:2889-2892, Kluwer Academic Publishers,
Dordrecht (1999).
[4]
SEQUENCE FROM N.A.
STRAIN=I1144;
MEDLINE=20031519; PubMed=10563807;
Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
Parot P., Vermeiglio A.;
"Dark aerobic growth conditions induce the synthesis of a high
midpoint potential cytochrome c8 in the photosynthetic bacterium
Rubrivivax gelatinosus";
Biochemistry 38:15238-15244(1999).
EMBL; AB034704; BAA94063.1; -;
PIR; T5091G; T5091G.
GO; GO:0016491; F:oxidoreductase activity; IEA.
GO; GO:0009058; P:biosynthesis; IEA.
GO; GO:0006118; P:electron transport; IEA.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac phytoene_dh.
InterPro; IPR000205; NAD_B5.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
SEQUENCE 511 AA; 56818 MW; 95648971960333903 CRC64;
Query Match 42.2%; Score 1099.5; DB 2; Length 511;
Best Local Similarity 42.3%; Pred. No. 1e-80;
Matches 204; Conservative 100; Mismatches 177; Indels 1; Gaps 1;
6 VIGAGFGGLALAIRLOAAGIPVLLSORDKPGGRAYVVOEQGTFDAGTIVTDSALIEE 65
23 VVSGFGGMAAVALAAKGRVTVLEKLDAPGRAYVHRRGHVFDAGTIVTVPLFDE 82
66 LFALAGKQLKDYVELLPVTPFYELCHWESKVFNYNDQALQAQIQENPRDVAGYRAFL 125
83 LWLAGRKFSDDIELKSLDPFYIRFDGDDHPDYGSDPARMAEVRRISSPSDAEGFERFM 142
126 DYSAVFNEGYKLGTVFPFLSPFKOMLRAAPOLAKQAMRSVSVKAGYIEDHLRQAFSP 185
143 READQCVELGFRITGLDKAFDTVGDLIAAPLIIRLGRWESLHQMVSSSHLPKPLAIA 202
186 HSLVAGNPPATSIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGEVVLNARVS 245
203 QSLILGNPFVSMYALYNALERQGVHMGVGGTGLIRGLVDVFEKGGMTRLKAEVK 262
246 HMETVGDKIQAQVLEGRFPETCAVASNADVVHTYDRLSQAAPAAKQAKKLOSKRMNS 305
263 RIEVDNGVATGVTADGERIPADIVVCNEDTGYLKNLVDARWRKHTDARIERHYSMG 322
306 LFVLYPGLNHHDDQLAHHTVCFGRPRYRELIHIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
323 LFVWYFGTDRRYEDVPHMMVLGPRYRELLDDIFRKKLAGDPSIYLRPTATDSMAPA 382
366 GCGSYVVLAPVPHLGTANLDMAVEGRRLDRIDFYLEQHYMPGLRSQVLRHMFPPDFR 425
383 GCDTFVAMPVPHLGSCT-DWTTQAEPRYSQVQEALEERTVLPGLQHLRVSCFTTPLDFQ 441
426 DELNACGSFAFVPEPILTOSAFRPHNRDKHIDNLYLVGACGTHPCAGIPGVIGSAGATAG 485
442 HRLSYKAGFGLEPGLLQSAVFRPHNRSEDVKNLFNMGASTHGAGVGVINGAKALES 501

496 EX 487
502 VL 503
RESULT 10
Q840T3 PRELIMINARY; PRT; 513 AA.
ID Q840T3;
AC Q840T3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CRII.
OS Rhodocyclus gelatinosus (Rhodospseudomonas gelatinosa).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Rubrivivax.
CX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SI;
RA Ouchane S., Steunou A.-S., Astier C.;
"Photosynthesis Gene Regulation in Rubrivivax gelatinosus:
RT Transcription Factor PpsR is Involved in both Negative and Positive
RT Control";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY234385; AAC93135.1; -;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene_dh.
DR InterPro; IPR000205; NAD_B5.
DR InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
SEQUENCE 513 AA; 57074 MW; 9C46447BB02F682D CRC64;
Query Match 42.0%; Score 1093.5; DB 2; Length 513;
Best Local Similarity 42.1%; Pred. No. 3.2e-80;
Matches 203; Conservative 101; Mismatches 177; Indels 1; Gaps 1;
6 VIGAGFGGLALAIRLOAAGIPVLLSORDKPGGRAYVVOEQGTFDAGTIVTDSALIEE 65
25 VVSGFGGMAAVALAAKGRVTVLEKLDAPGRAYVHRRGHVFDAGTIVTVPLFDE 84
66 LFALAGKQLKDYVELLPVTPFYELCHWESKVFNYNDQALQAQIQENPRDVAGYRAFL 125
85 LWLAGRKFSDDIELKSLDPFYIRFDGDDHPDYGSDPARMAEVRRISSPSDAEGFERFM 144
126 DYSAVFNEGYKLGTVFPFLSPFKOMLRAAPOLAKQAMRSVSVKAGYIEDHLRQAFSP 185
145 READQCVELGFRITGLDKAFDTVGDLIAAPLIIRLGRWESLHQMVSSSHLPKPLAIA 204
186 HSLVAGNPPATSIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGEVVLNARVS 245
205 QSLILGNPFVSMYALYNALERQGVHMGVGGTGLIRGLVDVFEKGGMTRLKAEVK 264
246 HMETVGDKIQAQVLEGRFPETCAVASNADVVHTYDRLSQAAPAAKQAKKLOSKRMNS 305
265 RIEVDNGVATGVTADGERIPADIVVCNEDTGYLKNLVDARWRKHTDARIERHYSMG 324
306 LFVLYPGLNHHDDQLAHHTVCFGRPRYRELIHIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
325 LFVWYFGTDRRYEDVPHMMVLGPRYRELLDDIFRKKLAGDPSIYLRPTATDSMAPA 384
366 GCGSYVVLAPVPHLGTANLDMAVEGRRLDRIDFYLEQHYMPGLRSQVLRHMFPPDFR 425
385 GCDTFVAMPVPHLGSCT-DWTTQAEPRYSQVQEALEERTVLPGLQHLRVSCFTTPLDFQ 443
426 DELNACGSFAFVPEPILTOSAFRPHNRDKHIDNLYLVGACGTHPCAGIPGVIGSAGATAG 485

b 444 HRLSVKAGAGLEPLLLQSAYPRHNSRSEDVKNLFWMGASTHPCAGVPGVMSAKALES 503
y 486 LM 487
b 504 VL 505
RESULT 11
9RW08
C Q9RM08 PRELIMINARY; PRT; 548 AA.
F 01-XAY-2000 (TREMBLrel. 13, Created)
F 01-XAY-2000 (TREMBLrel. 13, Last sequence update)
F 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
F Phytocene dehydrogenase.
X DR0861.
X Deinococcus radiodurans.
X Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
X Deinococcaceae; Deinococcus.
X NCBI_TaxID=1299;
X (1)
X SEQUENCE FROM N.A.
X STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
X MEDLINE=20036896; PubMed=10567266;
X White O., Eisen J.A., Heidelberg J.F., Hickey E.X., Peterson J.D.,
X Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
X Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
X Vamathevan J.J., Lam P., McDonald L., Unterback T., Zalewski C.,
X Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
X Ketchum K.A., Nelson K.B., Salzberg S., Smith H.O., Venter J.C.,
X Fraser C.M.;
X "Genome sequence of the radioresistant bacterium Deinococcus
X radiodurans R1."
X Science 285:1571-1577(1999).
X EMBL; AF001940; AF010439.1; -.
X PIR; C75466; C75466.
X TIGR; DR0861; -.
X GO; GO:0006118; P:electron transport; IEA.
X InterPro; IPR002937; Amino oxidase.
X InterPro; IPR008151; Phytin dehydro.
X Pfam; PF01593; Amino oxidase; 1.
X ProDom; PD139017; Phytin dehydro; 1.
X Complete proteome.
X SEQUENCE 548 AA; 60188 MW; 80A20729284A9A6C CRC64;

Query Match 40.3%; Score 1051; DB 16; Length 548;
Best Local Similarity 43.5%; Pred. No. 1e-76;
Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9;
2 KPTTVIGAGGGLALAIRLQAGIPVLLLEQDKPGRAYVYQEQGFTFDAGPTVITDP 60
17 KTAIVIGAGGGLALGIRLSIGFDTTILERLDGPGRAYQKRTPDGVVDMGPTVLTVP 76
61 SAIELEFA-----LAGKQKQ-----YVELLPVTPYRLCWES 93
77 HFIEELFAERDRAGLDADYPPEVLGSEYKEGVGGPHTSRVTLVPLPTFRIVFHD 136
94 GKVFYNDQQAQLBAQIQCFPRDVGVRAPLDYSRAVFNEGYLKLTGVPFLSKWLRA 153
137 GTFYDGDGPSTRRQFAELAPGLAGYERPHADABAIFRRGFLGLGYTHFGDVPTWLRV 196
154 APOLAKLQARSVSVKAGYIEDHLQARSFHLVGGNPPATSTYTLIHALERWGV 213
197 VFDLLKLDVATLFTSFSTSKYFSQSKLRQVTSFTLLVGGNPLSPALYAMIHFEVKTWGI 256
214 WFPGGGTGALVNGMKIKLFDGLGGBWLNARVSHMETVGD-----KIQAVOLEDGRFETC 268
257 HYAMGGTGAIVGLVQKEEELGGAIRGAGVDEVLVDGNIPLPKRTAGVRLSGEELRAD 316
269 AVASNADVVHTYRDLLOHAPAAAKKLOSKRMSN--SLFVLYFLGNHHD--QLAHHTV 325
317 LVASNGDWANTY--LKRVRPSARLVNSDLRVKAASEMSLLVYVYFGPGGDDLPKLNHI 374

QY 326 CFCPRYRELTHIEIFNHDGLARDFSLYHAPCVTDPSLAPEGCGSYVVLAPVPHLGATNLD 385
Db 375 LLGPRYELLSEIFGCTKRLGDFOSQLVHPTLTDPALAGHAAATLVVPVING-SGID 433
QY 386 WAVEGPRLRDRIPDYLEQH-YMFGRLRSQVLTHERMTTFDFRDELNAWQGSASFVPILTO 444
Db 434 WDVEGPKLAALADIERRGLIFGLRERLTHFEFTPDYFAGTLDLSYLGNAFGPBPRLVQ 493
QY 445 SAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLD 490
Db 494 SAFFRPHNRSEDLNHFYLVGAGAQPGAGTSPVMSAKMTARLIAED 539
RESULT 12
Q8KZ28
ID Q8KZ28 PRELIMINARY; PRT; 514 AA.
AC Q8KZ28;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
Db Phytocene dehydrogenase.
GN CRTI.
OS uncultured proteobacterium.
OC Bacteria; Proteobacteria; environmental samples.
OX NCBI_TaxID=153809;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21822632; PubMed=11832943;
RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs."
RL Nature 415:630-633(2002).
DR EMBL; AF008921; AF048646.1; -.
DR GO; GO:0018491; F:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytocene_dh.
DR InterPro; IPR000205; NAD_B5.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytin dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 514 AA; 56824 MW; D5177500CAB56773 CRC64;

Query Match 37.8%; Score 984.5; DB 2; Length 514;
Best Local Similarity 41.8%; Pred. No. 2.4e-71;
Matches 206; Conservative 77; Mismatches 191; Indels 19; Gaps 5;
6 VIGAGGGLALAIRLQAGIPVLLLEQDKPGRAYVYQEQGFTFDAGPTVITDPSAIEE 65
15 VIGAGGGLAAMRLGAKGYAVTVLDKLDVGGSGSVTQDGHFGLGPTIVTPKVFES 74
66 LPALAGKLDYVLELTPVTPYRLCWESKGVNTDNOAQLAQIQCFNPRDVAGYRAFL 125
75 LWAACGRDFHADVDLRPLEPFYTRWPDGYSFRASGDDKMQSEVQRINPADLPGRYRFL 134
126 DYSRAVFNEGYLKLTGVPFLSKWLRAAPQLAKLQARSVSVKAGYIRDEHLROAFSP 185
135 KDSQRYIIIGEGWVAPMHLWTLLKVLPTFAMLRADRSIYGLAARRVXDERLMLSP 194
186 HSLVGGNPPATSTYTLIHALERWGVPPRGOTGALVNGMTKLFQDLGGEVVLNARVS 245
195 HPLFTGGDPHVTSTYALVAHLEKTYGVHVGWGVQIQIADAMAARVRAQGGQI----- 247
246 HMETVGDKI-----QAVQLDGRREFTCAVASNADVHTYRDLLOSHPAAKQAKIKQ 298
248 HONAVADEILLINGAQAVALTQGRFDADPLTVSNADAGTYDHLNHRERRRRTTKLA 307
299 SKRMSNLSFLYFLGNHHDQLA---EHTVCFGPYRELTHIEIFNHDGLAEDFSLYLHAP 355
308 RKRWSMGLFVWYFGTRGTAGRWADVGHHTIANGPRYKGLLRDIFLKGRLSDMSLYIHRP 367


```
2Y 356 CTTDPSLAPEGCGSYVLAIPVPHLGTAN-LDNAVEGPRRLDRIDFVLEQHYMGLRSQLV 414
DB 368 SVTDPSPVAPAGDDTFYVLSPLGKKNKWDQKEMP:YFAKAAAEVK-LMPGFACIS 426
QY 415 THRTFTPDFFDELNAWOGSAFSEVPEILTQSANFRPHNRKHDIDNLYLVGAGTHPGAGIP 474
DB 427 TETFTTPTTFEDVLSPHGAGFSIEPRILQSANFRPNVSEEARGLYLVGAGTHPGAGLP 486
QY 475 GVIGSAKATAGIM 487
DB 487 GVISAASVLSKLV 499

RESULT 13
292G56 PRELIMINARY; PRT; 517 AA.
AC Q9ZGE4;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Diapophytoene dehydrogenase CrtN.
SN CrtN.
SS Helicobacillus mobilis.
SC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
SC Helicobacillus.
XX NCBI_TaxID=28064;
XX [1]
XX SEQUENCE FROM N.A.
XX MEDLINE=93061957; PubMed=9843979;
XX Xiong J., Inoue K., Bauer C.E.;
XX "Tracking molecular evolution of photosynthesis by characterization of
XX a major photosynthesis gene cluster from Helicobacillus mobilis.";
XX Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
XX EMBL; AF080002; AAC84034.1; -.
XX PIR; T31463; T31463.
XX GO; GO:0016491; 2-oxido-reductase activity; IEA.
XX GO; GO:0006118; P:electron transport; IEA.
XX InterPro; IPR000759; Adrnx reductase.
XX InterPro; IPR002937; Amino Oxidase.
XX InterPro; IPR008151; Phytin dehydro.
XX InterPro; IPR001100; Pyr redox.
XX Pfam; PF01593; Amino oxidase; 1.
XX PRINTS; PR00419; ADXRDTASE.
XX PRINTS; PR00411; ENDRDTASE1.
XX ProDom; PD139017; Phytin dehydro; 1.
SQ SEQUENCE 517 AA; 58340 MW; BB933977999C4587 CRC64;

Query Match 28.9%; Score 752; DB 2; Length 517;
Best Local Similarity 34.2%; Pred. No. 2e-52;
Matches 170; Conservative 91; Mismatches 214; Indels 22; Gaps 7;

2Y 6 VIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQGGFTPDAGPTVITDPSAIEE 65
DB 13 VVSGAGGSAVAELANQGDVTLREKATPGGLSAIQAGYSIDVGPPTIMMNDVPHQ 72
QY 66 LPALAGKQKDYVELLPVTPFPYRLCHWESKGVNYDNDQALEAQIQPNPRVAGTRAF 125
DB 73 YFKDGRNIEDYLDLVANPCYHLHFTDGTGCMKPSIDLKELDEIRSFNPDVDVGYRL 132
QY 126 -----DYSRAVNEGYLKLGTVPFLSFK-DMLRAAPQAKIQAVRSVYKAVGIEDE 177
DB 133 AQIHRRVQVAKEKIE---KSFTKPSDFNIDTLGWMQLRTL---NNMYDDIARFIKDB 186
QY 178 HLRAQFPHSLVCGNPPATSSYITLI-----HALREWGVPFRPGGTGALVNGIMIKLFOD 233
DB 187 RLRLSLTFQAIYLVGSPDPAPSITLYIGVVEHGLS---GWVYKGGNAITQALVKLLGE 243
QY 234 LGGSHVNLARVSHMETVGDQIAQVQLEDGRFRFETCAVSNADVHTYRDLSSQHPAAAKQ 293
DB 244 FGGSLRVNAEVEQILIQGAVGVRLANGEVLKADVVISNADFPYTWENLVPAHSGKXT 303
QY 294 AKKIQSKRMNSLFLVYFGLNHHHDQLAHHTVCGPRYRELIIHEIFNHDGLAEDFSLYLH 353
```

```
DB 304 POKLNSMEVTCGAFMLYLGVNRRYDNLHLNIIYTPDYKYSMDLELFTROQLPQDFAMIVY 363
QY 354 APCVTDPSLAPEGCGSYVLAIPVPHLGTANLDAVEGPRRLDRIDFVLEQHYMGLRSQ 413
DB 364 RPTKYDSDVAPGKDIYLVLPVNL-SSGIDWKKEKTHRYRELIVKGLERQGVTDLSKHI 422
QY 414 VTHRTFTPDFFDELNAWOGSAFSEVPEILTQSANFRPHNRKHDIDNLYLVGAGTHPGAGI 473
DB 423 EFERIYTPETTFQNRFNYYQGAAGLAPSLFQSGYFRPHIKSEKVPNLYFSGASVHPGGV 482
QY 474 PGVIGSAKATAGLMLD 490
DB 483 PVVLVCGKLWSEQVLAD 499

RESULT 14
Q9ENMO PRELIMINARY; PRT; 511 AA.
AC Q9ENMO;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase (phytone desaturase) (EC 1.3.-.-).
SN OB2461.
SS Oceanobacillus thevensis.
SC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
XX NCBI_TaxID=182710;
XX [1]
XX SEQUENCE FROM N.A.
XX STRAIN=HT831 / DSM 14371 / JCM 11309;
XX MEDLINE=22220767; PubMed=12235376;
XX Takami H., Takaki Y., Uchiyama I.;
XX "Genome sequence of Oceanobacillus thevensis isolated from the Iheya
XX Ridge and its unexpected adaptive capabilities to extreme
XX environments.";
XX Nucleic Acids Res. 30:3927-3935(2002).
XX EMBL; AP004601; BAC14417.1; -.
XX GO; GO:0016491; P:oxidoreductase activity; IEA.
XX GO; GO:0006118; P:electron transport; IEA.
XX InterPro; IPR000759; Adrnx reductase.
XX InterPro; IPR002937; Amino Oxidase.
XX InterPro; IPR000205; NAD B5.
XX Pfam; PF01593; Amino oxidase; 1.
XX PRINTS; PR00419; ADXRDTASE.
XX Oxidoreductase; Complete proteome.
SQ SEQUENCE 511 AA; 58463 MW; 8E70FD9F71D406AB CRC64;

Query Match 27.1%; Score 706.5; DB 16; Length 511;
Best Local Similarity 33.2%; Pred. No. 9.6e-49;
Matches 170; Conservative 91; Mismatches 212; Indels 39; Gaps 12;

QY 2 KPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQGGFTPDAGPTVITDPS 61
DB 4 KKVIVIGAGVAGLASAIRLQAGFNVEYKESMPGKMHQITKDGTFDLPGLSIVMPE 63
QY 62 AIELFALAGKQKDYVELLPVTPFPYRLCHWESKGVNY--DNDQALEAQIQPNPRDVA 119
DB 64 LYRAVFELTGRNPDYIPMERIDPMNVNFFGNTPEDRYQISSLIQIDIEIAISKDAA 123
QY 120 GYRAFLD--YSRAVNEGYLKLGTVPFLSFKDLRAAPQAKLQ-----AMRSVYSKVA 171
DB 124 GFITYLNKIYOR-----FQIAKDHFIQRFPRHPQWQFFPKTLWNLGLKRTL 169
QY 172 G-----YIEDHLRQAFSFLSLVGGNPPATSSYITLIHALREWGVPFRPGGTGAL 223
DB 170 GNADNFITKVKIKDRLRQMSIFQTLXIGISPYNGPSLYTIIPMEMLYGVWFKGGMYTM 229
QY 224 VNGMIKLFQDLGSEWLNARVSHMETVGDQIAQVQLEDGRFRFETCAVSNADVHTYRDL 283
DB 230 ATSEKELFLELGGTIHYNVSPKVIIEKKATGIKLEN-LEINADYVVCNADFPYSKML 288
QY 284 LSQHPAAAKQA-KKLOSKRMNSLFLVYFGLNHHHDQLAH-HTVCGPRYRELIIHEIFNH 341
```

289 IQDKAKGKYTDEKIDKMKYSCSCFIMYLGMNKYDEVSNVHNPFFSENKETMDIDFTG 348
342 DGLAEDESLYHAPCVTPDSPSLAPGCGSYVYLAVPPLHGTANLDMAVEG-PRLRDRIFDY 400
349 KRL-EDAFYTYIGSKYDPSLAPGKGIYVLPVSDLSAQYSGWDETAYYREKVEK 407
401 LEQ-HYMPGLASQIVTHMFTFPFRDELNAWQGSFVEPILTQSAWFRPHNRDKHIDN 459
408 LDE-DBFKGEKDEIVSETYMTPLDFESKFAYNACFLAPTLTSSQNHLPQSKAKNEN 467
460 LYLVGAGTHPCAGIPGVIGSAK-ATAGLMLEDL 490
468 LYFGSSHTPGAGVFIYLLSAKIATEELLKDD 499

SULT 15

UF87

Q7UF87 PRELIMINARY; PRT; 537 AA.

01-OCT-2003 (TRENBLrel. 25, Created)

01-OCT-2003 (TRENBLrel. 25, Last sequence update)

01-OCT-2003 (TRENBLrel. 25, Last annotation update)

Phytoene dehydrogenase (EC 1.14.99.-).

CRTI OR RB10370.

Rhodospirillum rubrum.

Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;

Planctomycetaceae; Fireillula.

NCBI_TaxID=117;

[1]

SEQUENCE FROM N.A.

STRAIN=1;

MEDLINE=22735913; PubMed=12835416;

Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,

Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,

Schlesner H., Amann R., Reinhardt R.;

"Complete genome sequence of the marine planctomycete Fireillula sp.

strain 1.";

Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).

EMBL; BX294151; CAD78796.1; -

Oxidoreductase; Complete proteome.

SEQUENCE 537 AA; 60136 MW; A16C389C746949B6 CRC64;

Query Match

Best Local Similarity 26.6%; Score 693.5; DB 16; Length 537;

Matches 162; Conservative 115; Mismatches 204; Indels 17; Gaps 9;

2 KPTTVIGAGFCGLAIARLQAGIPVLLRQKPGCGRAVYVQGGFTPDAGTPTIDPS 61

25 KKVWVGAGPGGLASAKQLAAGCGDVTILERRGQVGGRTSAIEIDGFRDCGPTFFLYPR 84

62 AIELFALAGKQKDYVELLPVTPFRLCWESGKVFVNDQALQLEAOIQCFNPRDVAGY 121

85 VLDRIFHSTGRDLMEVPMBELDPQYRLTGGGQQLDCTFDMEMBRQIAQFSQDVQQL 144

122 RAFLDYSAVFNKGLGTVPFLSKDKMLRAAPQL---AK-LQANRSVYKVGAGYIEDE 177

145 KRYMDNR-IKLERFPILESPPHSALDVMK--PSLGAAGKHLHPFRTLGKELERYFSOP 201

178 HLQAFSPHSLVCGSNPATTSIYTLHALERGVWVFRGGTCALVNGMIKLPQDLGG 237

202 RLVTAFQSKYLGNSPNCPSLSILSFLEYEGVPHPIGGCSRVSEKVAETAEEMGVK 261

238 VVLNARVSHMETVGDKIQAQLEDEGRFETCAVASNADVVVHTYRDLSQH-PAAAKQA-- 294

262 IRLNEPVDSEMEGRVRALHTQ-ADRYDADA FVWNAD---FADWTKVTPNASKRWS 316

295 -KKLQSKMSNSLFLVYGLNHHQDLAHITVCGPRYRLIHEIFNHDLAEDFSLYLH 353

317 DEQAKKFKSCSTVLYLHIEGLYEDUPHSHIHSIDYNRLREIETDHLISQDPSVYQ 376

354 APCVTPSLAPGCGSYVYLAVPPLHGTANLDMAVEGPRLRDRIFDYLEQHYMPGLRSOL 413

Db 377 NAGVTDPTLAPAGHSSLYLVVPVTH-DTENVDWSKGAAGPRELTLDKLGELGLTDVRDRI 435
QY 414 VTHRMFTPDFRDELNAWQGSFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPCAGI 473
Db 436 RVEHQITPDHQQSDYSYKGGATFNLAHNLGQMLHCRPRNRPEELDGYLVGCGGTHPSSEL 495
QY 474 PGVIGSAKATAGLMLEDL 491
Db 496 PVIVESSRISRLLLIQDL 513

Search completed: February 29, 2004, 14:51:01

Job time : 43.4556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 8.29213 Seconds
(without alignments)
3089.496 Million cell updates/sec

tle: US-09-941-947A-32

fect score: 2606

quence: 1 MKPTVIGAGGGLAAILR.....IPGVIGSAKATAGLMLEDLI 492

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

al number of hits satisfying chosen parameters: 141681

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2438	93.6	492	1 CRTI_PANAN	P21685 pantoea ana
2	2023	77.6	492	1 CRTI_ERWHE	P22871 erwinia her
3	1638	62.9	501	1 CRTI_AGRAU	P54978 agrobacteri
4	1007.5	38.7	524	1 CRTI_RHOCA	P17054 rhodobacter
5	959.5	36.8	518	1 CRTI_RHOSH	P54980 rhodobacter
6	757.5	29.1	517	1 CRTI_MYXXA	P54979 myxococcus
7	672.5	25.8	507	1 CRTI_STRGR	P54961 streptomyce
8	660.5	25.3	508	1 CRTI_STRSE	P54971 streptomyce
9	635	24.4	595	1 CRTI_NEUCR	P21334 neurospora
10	616	23.6	529	1 CRTI_MYXXA	Q02861 myxococcus
11	595.5	22.9	583	1 CRTI_MYXBL	P54982 phycococcy
12	530	20.3	621	1 CRTI_CERNC	P48537 cercospora
13	430.5	16.5	495	1 CRTD_RHOSH	Q01671 rhodobacter
14	369.5	14.2	494	1 CRTD_RHOCA	P17059 rhodobacter
15	171.5	6.6	469	1 P49_STRLI	P06108 streptomyce
16	140	5.4	527	1 AOFA_HUMAN	P21397 homo sapien
17	140	5.4	749	1 TR2M_AGRHR	Q09109 agrobacteri
18	136.5	5.2	570	1 CRTI_SOYBN	P28553 glycine max
19	132.5	5.1	527	1 PROX_PROFF	Q32434 propionibac
20	127.5	4.9	583	1 CRTI_LYCES	P28554 lycopersico
21	124.5	4.8	495	1 AOFPN_ASPNG	P46882 aspergillus
22	123.5	4.7	516	1 OKLA_AGRGH	P81382 agkistrodon
23	121.5	4.7	571	1 CRTI_MALZE	P49086 zea mays (m
24	121	4.6	527	1 AOFA_CANFA	P58027 canis fami
25	120.5	4.6	454	1 AOFPN_MYCTU	Q53320 mycobacteri
26	116	4.5	526	1 AOFA_RAT	P21398 rattus norv
27	116	4.5	657	1 YH09_RALSO	QBXP79 ralstonia s
28	115	4.4	508	1 FMSI_YEAST	P20284 saccharomyc
29	114.5	4.4	557	1 TR2M_PSSSS	P06617 pseudomonas
30	113.5	4.4	566	1 CRTI_ORYSA	Q92709 oryza sativ
31	113	4.3	582	1 CRTI_CAPAN	P80093 capsicum an
32	112	4.3	527	1 AOFA_BOVIN	P21398 bos taurus
33	111.5	4.3	417	1 YAAB_RHISN	P55349 rhizobium s

34	110	4.2	519	1 AOFP_RAT	P19643 rattus norv
35	110	4.2	567	1 FIG1_HUMAN	Q961q9 homo sapien
36	110	4.2	999	1 OXRP_RAT	Q63617 rattus norv
37	109.5	4.2	444	1 YGAF_ECOLI	P37339 escherichia
38	109.5	4.2	535	1 Y897_MYCTU	Q10555 mycobacteri
39	108.5	4.2	566	1 CRTI_AATH	Q07356 arabidopsis
40	108	4.1	519	1 AOFP_HUMAN	P27338 homo sapien
41	107.5	4.1	471	1 PPOX_MYXXA	P56601 myxococcus
42	106	4.1	519	1 AOFP_CAVPO	P58028 cavia porce
43	106	4.1	755	1 TR2N_AGRVI	P25017 agrobacteri
44	105.5	4.0	999	1 OXRP_HUMAN	Q9V411 homo sapien
45	104.5	4.0	999	1 OXRP_CRIGR	Q60432 cricetus

ALIGNMENTS

RESULT 1

ID	CRTI_PANAN	STANDARD;	PRT;	492 AA.
AC	P21685;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).			
GN	CRTI..			
OS	Pantoea ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoea.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
RX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	Coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
CC	!- FUNCTION: This enzyme converts phytoene into lycopene via the			
CC	intermediaries of phytofluene, zeta-carotene and neurosporene by			
CC	the introduction of four double bonds.			
CC	!- COPACTOR: FAD (Probable).			
CC	!- PATHWAY: Carotenoid biosynthesis			
CC	!- SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D90087; BAA14127.1; ..			
DR	PIR; D37802; D37802.			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac phytoene dh.			
DR	InterPro; IPR008151; Phytan dehydro.			
DR	Fam; F01593; Amino oxidase; 1.			
DR	ProDom; P0139017; Phytan dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE DH; 1.			
KW	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.			
FT	NP_BIND 5 38 FAD (ADP PART) (POTENTIAL).			
SQ	SEQUENCE 492 AA; 55007 MW; PA4CD4E34A9C6413 CRC64;			

Query Match 93.6%; Score 2438; DB 1; Length 492;
Best Local Similarity 93.5%; Pred. No. 3.8e-190;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKPTVIGAGGGLAAILRQAAGIPVLLLEQRDKPGRAVYVQEQGTFDAGPTVITDP 60

1 MKPTTVIGAGGGIALAIRLQAAGIPVLLLEQDKPGRAVYVEDQGTFTDAGPTVITDP 60
61 SAIEELFALAGKOLKOVYELLPTVPFVRLCWSSGKVFYNDQOALQLEAQIQGFNPRDVAG 120
61 SAIEELFALAGKOLKOVYELLPTVPFVRLCWSSGKVFYNDQOALQLEAQIQGFNPRDVAG 120
121 YRAPLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKIQAWSYVSKVAGYIIDEHLR 180
121 YRAPLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKIQAWSYVSKVAGYIIDEHLR 180
181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKQASK 300
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKQASK 300
301 RMNSLFLVFLGNNHHDQIAHVTGCGPRYRELIIHIFNHDGLAEDPSLYLHAPCVITDP 360
301 RMNSLFLVFLGNNHHDQIAHVTGCGPRYRELIIHIFNHDGLAEDPSLYLHAPCVITDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 2

RTI ERWHE
D - CRTI ERWHE STANDARD; PRT; 492 AA.
C P22871;
T 01-AUG-1991 (Rel. 19, Created)
T 01-AUG-1991 (Rel. 19, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CRTI.
S Erwinia herbicola.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=549;
X [1]
P SEQUENCE FROM N.A.
C STRAIN=EHO10;
X MEDLINE=9108634; PubMed=2263648;
X Armstrong G.A., Alberti M., Hearst J.E.;
T "Conserved enzymes mediate the early reactions of carotenoid
T biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
T Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979 (1990).
C -!- FUNCTION: This enzyme converts phytoene into lycopene via the
C intermediaries of phycofluene, zeta-carotene and neurosporene by
C the introduction of four double bonds.
C -!- COFACTOR: FAD (Probable).
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C
R EMBL; M38423; AAA24820.1; -;
R EMBL; M87280; AAA64981.1; -;

DR A39273; A33120.
DR InterPro; IPR02937; Amino oxidase.
DR InterPro; IPR008150; Bac.phytoene dh.
DR InterPro; IPR008151; Phyto_dehydro.
DR Pfam; PF01593; Amino_oxidase; 1.
DR PROSITE; PD139017; Phyto_dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 38 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;

Query Match 77.6%; Score 2023; DB 1; Length 492;

Best Local Similarity 76.8%; Pred. No. 1.8e-156; Indels 0; Gaps 0;

Matches 377; Conservative 42; Mismatches 72; Indels 0; Gaps 0;

QY 1 MKPTTVIGAGGGIALAIRLQAAGIPVLLLEQDKPGRAVYVEDQGTFTDAGPTVITDP 60
DB 1 MKPTTVIGAGGGIALAIRLQAAGIPVLLLEQDKPGRAVYVEDQGTFTDAGPTVITDP 60
QY 61 SAIEELFALAGKOLKOVYELLPTVPFVRLCWSSGKVFYNDQOALQLEAQIQGFNPRDVAG 120
DB 61 SAIEELFALAGKOLKOVYELLPTVPFVRLCWSSGKVFYNDQOALQLEAQIQGFNPRDVAG 120
QY 121 YRAPLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKIQAWSYVSKVAGYIIDEHLR 180
DB 121 YRAPLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKIQAWSYVSKVAGYIIDEHLR 180
QY 181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKQASK 300
DB 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKQASK 300
QY 301 RMNSLFLVFLGNNHHDQIAHVTGCGPRYRELIIHIFNHDGLAEDPSLYLHAPCVITDP 360
DB 301 RMNSLFLVFLGNNHHDQIAHVTGCGPRYRELIIHIFNHDGLAEDPSLYLHAPCVITDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
QY 421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDLI 491
DB 481 KATAGLMLEDLI 491

RESULT 3

CRTI_AGRAU
ID CRTI_AGRAU STANDARD; PRT; 501 AA.
AC P54978;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96052243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
RT biosynthesis gene cluster and astaxanthin biosynthetic pathway
RT proposed at the gene level.";

J. Bacteriol. 177:6575-6584 (1995).

-!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediaries of phytylfene, zeta-carotene and neurosporene by the introduction of four double bonds (By similarity).

-!- COFACTOR: FAD (Probable).

-!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin biosynthetic pathway.

-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D58420; BAA9594.1; -
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008150; Bac phytoene dh.
 Pfam; PF01593; Amino oxidase; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE DH; 1.
 Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SEQUENCE 501 AA; 54806 MW; 5P251AF11D679358 CRC64;

Query Match 62.9%; Score 1638; DB 1; Length 501;
 Best Local Similarity 63.9%; Pred. No. 3.3e-125;
 Matches 313; Conservative 61; Mismatches 116; Indels 0; Gaps 0;

2 KPTTVIGAGFGGLALAIQLQAGIPVLLLEQRKPGGRAYVYQGGTFDAGPTVIDPS 61
 9 KTAIVIGAGFGGLALAIQLQAGIATTLVEARDKPGGRAYVYHDDGHLFDAGPTVIDPD 68
 62 ALEELFALAGKQDKDYVELLPVTFYRLCWSEKGVFNFDNDQALENQLIQFNPDRVAGY 121
 69 ALKELWALTQDMARDVTLMPVSPFYRLMPPGKGVFDVNEADQLERQIAQFNPDDLEGY 128
 122 RAFLDYSRAVFNKGLGTVPFFLSFKDMLRAAPQLAKLQAWRSVSKVAGYTEDHLRQ 181
 129 RFRDIYAEVYQEGYKLGTVTFPFLKGLQMLKAAALMKLEAYKSHAKVATFIKDPVLRQ 198
 182 AFSFSLVLGNGNPATSSYITLIHALREMGVPPRGGTGALVNGXIKLPDLGGEVVLN 241
 189 AFSYHTLLVGNPFTSSYIALNHALERRGGVWPAKGTNQLVAGVWALFERLGQMLN 248
 242 REVSHMETVGDKIQAQVLEDGRRFETCAVSNADVHTYRLLSQHPAAAKQAKKLSKR 301
 249 AKVARIDTGRATGVTIADGRALTADNVASNGVDMVNYRLLIGHTARGOSRAKSLNAKR 308
 302 MSNSLFLVLYFGLNHHHDLAHTTVCFGPRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPS 361
 309 WMSLFLVLFHGLREAPKDVAAHTILFGPRYKELVNEIFKPKLAEDPSLYLESPTDPE 368
 362 LAPEGCGSYVYLAIPHGHGTANLWAVEGPIRLDRIDYLEQHYMPGIRLSQVTHRMFTF 421
 369 MAPGMSHTYVYLAIPHGHGRADIDWAVEGPIYADRIILASLEERLIPLNLANLTTRIFTP 428
 422 PDRFELNAWGSAPSVPEPILTSQAWFPHNRDGHIDNLYLVAGTHPGAGIPGVIGSAK 481
 429 SDFASELNAHGSAPSVPEPILTSQAWFPHNRDXTIRNFVLVAGTHPGAGIPGVIGSAK 488
 482 ATAGLMLEDL 491
 489 ATAQWMLSDL 498

.SULT 4

TI_RHOCA
 CRTI_RHOCA
 P17054;

PRT: 524 AA.

STANDARD;

Query Match 38.7%; Score 1007.5; DB 1; Length 524;
 Best Local Similarity 41.5%; Pred. No. 5.1e-74;
 Matches 204; Conservative 89; Mismatches 194; Indels 5; Gaps 3;

DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 GN CRTI.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
 OC Rhodobacteraceae; Rhodobacter.
 CX NCBI_TaxID=1061;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / St Louis, and BEC404;
 RX MEDLINE=89313663; PubMed=2747617;
 RA Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
 RT "Nucleotide sequence, organization, and nature of the protein
 products of the carotenoid biosynthesis gene cluster of Rhodobacter
 capsulatus.";
 RT J. Biol. Chem. 264:13109-13113 (1989).
 RN [3]
 RN ERRATUM.
 RA Bartley G.E., Scolnik P.A.;
 RL J. Biol. Chem. 264:18260-18260 (1989).
 RN [4]
 RN SIMILARITY TO CAROTENOID DESATURASES.
 RX MEDLINE=90368827; PubMed=2144293;
 RA Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
 RT "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
 crassa are structurally and functionally conserved and contain
 domains homologous to flavoprotein disulfide oxidoreductases.";
 RT J. Biol. Chem. 265:16020-16024 (1990).
 CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
 intermediary of phytylfene by the symmetrical introduction of two
 double bonds at the C-11 and C-11' positions of phytoene.
 CC -!- COFACTOR: FAD (Probable).
 CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; J04969; AAA50313.1; -
 DR EMBL; X52291; CAA36533.1; -
 DR EMBL; Z11165; CAA77540.1; -
 DR PIR; A32617; A32617.
 DR InterPro; IPR000759; Admrx_reductase.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac phytoene dh.
 DR InterPro; IPR000205; NAD BS.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino oxidase; 1.
 DR PRINTS; PR00419; ADXEDTASE.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE DH; 1.
 KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 KW Oxidoreductase; FAD; Flavoprotein; NAD.
 FT NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 524 AA; 57978 MW; 6425A7E5A06A6B9 CRC64;


```
422 EDRDE-LANAGQSAFSEVPILTQSAFPRPNRDKHIDNLYVAGTHPGAGTGGVIGSA 480
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 ETEFRYLSPW-GAGFSEIPILOSAFPRPNRDKHIDNLYVAGTHPGAGTGGVIGSA 492
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
481 KATAGL 486
: : : : :
493 EWAKL 498

SUET 6
TU MYXXA
CRUJ MYXXA STANDARD; PRT; 517 AA.
P54979;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
CARAZ.
Mycococcus xanthus.
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Cystobacterineae; Myxococcaceae; Myxococcus.
NCBI_TaxID=34;
[1]
SEQUENCE FROM N.A.
STRAIN=DKJ050;
MEDLINE=96061955; PubMed=7588751;
Botella J.A., Murillo F.J., Ruiz-Vazquez R.M.;
"A cluster of structural and regulatory genes for light-induced
carotenogenesis in Myxococcus xanthus.";
Eur. J. Biochem. 233:238-248(1995).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z21955; CRA79956.1; -
PIR; S32169; S32169.
InterPro; IPR000759; Adrndx_reductase.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR000205; NAD_BS.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
PRINTS; PR00419; ADXRDTASE.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 11 44 FAD (ADP PART) (POTENTIAL).
SEQUENCE 517 AA; 57783 MW; CSCDD74AB9F9F8CB CRC64;

Query Match 29.1%; Score 757.5; DB 1; Length 517;
Best Local Similarity 35.6%; Pred. No. 9.8e-54;
Matches 173; Conservative 88; Mismatches 216; Indels 9; Gaps 6;

6 VVGAGFGGLAIALRQAGTIPVLLRDRKPGGRAYVYQCGTTFDAGPTVTDPSAISE 65
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
12 VVGAGVGGLAAARLARGQFDVQVFEKTOGPGGRCNRLQVDGTTDLGPTVILMPVFEE 71
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 LFALAGKOLXDYVELLPVTPPYRLCWESGKVENVDNDQALEAQIQGFNPRDVAGYRAFL 125
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
72 TFRVGRRIEDYILLRCDDPNRVHFRDRSDVFTTSELCHAGRELKRVGSGVARYLAF 131
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 DYSAVFNEGYKLGTVFFLISFDKML--RAAPLAKLQAWRSVSKVAGYIEDEHLRQAP 183
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 132 AGRVQVYETSLDHLVGRNYAGLRDYLSPRVLARLQVFAHRMTADVERFFQDERLRAM 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 SFHSLVGGNPFATSSYITLHALERBNGWYFPRGGTCAALVNGMIKLPQDLGGEVVLNAR 243
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 TFOYWLGVSPVSPAVYGLLPFTTGLGVWFKGGLVYAFQALERLARERBGRVFRHYGAP 251
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 VSEMETVGDKIOAVOLEDRPFETCANASNADVHTYRDLLSQHPAAAKQAKLQSKGMS 303
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 VERILTDGGRTRGVRLGEGVEVDADVLGNADLPYAYEKLDPKATTLKRKEKL---RYT 308
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 NSLFVLYFELNHHEDQLAHTVCFGPYRELIFHNDGLAEDEFSLYHAPCVTDPSLA 363
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 SSGYMLYLGKRRYPPELLHNHVFGROYKGSFDDIFEF-RVPEDPSFYVNAFTETDASLA 367
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 364 PEGCGSYVILAPVPHLGTANLDAVEGPRLDRIDYLEQHYMGLASQL-VTHRMETPF 422
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 PEGKDALYLVLPVPH-QEPDLMDKVEGPKVRAKFAFARVABELGFFSLEDSEVERRSSTPD 426
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 DFRDELNAWQGSFSEVPEILTQSAWFRPNRDKHIDNLYVAGTHPGAGTGGVIGSAK- 481
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 DWAGTFNLARGSGFGLSQNFQIGFFRPSNQDARVKNLFFVGASTQGTGLPTVLISARL 486
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 482 ATAGLM 487
: : : : :
Db 487 VTERLM 492
: : : : :

RESULT 7
CRUJ_STRGR STANDARD; PRT; 507 AA.
AC P54981; P72447;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI OR CRTPE.
OS Streptomyces griseus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyceinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JA3933;
RX MEDLINE=97074881; PubMed=9917308;
RA Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
RT "Activation and analysis of cryptic crt genes for carotenoid
biosynthesis from Streptomyces griseus.";
Mol. Gen. Genet. 252:658-666(1996).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; L37405; AAA91950.1; -
DR EMBL; X95596; CAA64850.1; -
DR InterPro; IPR000759; Adrndx_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac_phytoene_dh.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.

```


R PRODOM; PD139017; Phytin dehydro; 1.
 R PROSITE; PS00982; PHYTOENE DH; 1.
 W Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 T NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 Q SEQUENCE 507 AA; 5450 MW; FBB57F7FE696B2AC CRC64;

Query Match 25.8%; Score 672.5; DB 1; Length 507;
 Best Local Similarity 35.0%; Pred. No. 7.6e-47;
 Matches 171; Conservative 80; Mismatches 214; Indels 10;

Y 6 VIGAGGGLALALRLQAAGIPVLLLBQRDKPGRAVYVQEGFTFDAGTWTIDPSAIEE 65
 b 13 VVGAGLAGLAAHLHLCAGRRVTVVEREDVPGGRAGLLESGFRIDTGTVTWIDVED 72
 Y 66 LFALAGKQLKDYVELLPVTPFYELCWESKVFYNDQALQLEAQIQFNPDRVAGYAF 125
 b 73 AFAVGERMADRLRLAPAYRAREADGSLDVTDTGAAMEAAVEEF-----AGARQAV 127
 Y 126 DYSR-AVFREGLKLTVPPL--SFQMLRAA-PQAKLQA---WRSVYSKVAGYIEDEH 178
 b 128 GYRLRLIWLRLYRVQMRRTDNTFSPQLAHPDLARLAALGGFRDLARIGHVFSDER 187
 Y 179 LRQAFSPHLLVGNPNFATSSIIYTLHALERWGVNFPFGGTGALVNGMKLFQDLGGEV 238
 b 188 LRVFSPQALYAGVPPARALAAVAVIYMDTVAGVYFPRGGMHALPRAMADAASDAGAVL 247
 Y 239 VLNRVSHMETVGKQAVOLEGRRPETCAVASNADVHTTRDILLSQHPAAKQAKLQ 298
 b 248 RYGVTRFLRSRGRVTAV-VTDQRIADAVVLTDPDLPVSYR-LLGRTP-----HRPL 299
 Y 299 SKMNSLFLVYFGLNHHDDLAHTVCFGRYRELHIEIFNHGDLAEFSLYLHAPCVT 358
 b 300 PLRHSFSAVILHGTDTWDLAHTTISFGAAWNTFHELTTRGLMSDPSLLITPTAT 359
 Y 359 DPLSLAPGCGSYVYVLAPEVH--LGTANLDWAVEGPRRLDRIDFYLEQHYNPGRLSOLVTH 416
 b 360 DPLSLAPGKHLHYVLAPEVH--LGTANLDWAVEGPRRLDRIDFYLEQHYNPGRLSOLVTH 419
 Y 417 RMTPEPFRDELNAWQSSAFSVEPILTQSAWFRPHNRKHIDNLYLVGAGTHPGAGICV 476
 b 420 GLVTPDVTWTAQGH-AGTPFSVAHTFTQTPFRNLRVGTNVNAVLAGCGTTTGGVGVTV 478
 Y 477 IGSAKATA 484
 b 479 LISGLKAA 486

RESULT 8

RTI STRSE STANDARD; PRT; 508 AA.
 AC P54971;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 CR1.
 NC Streptomyces setonii.
 NC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 NC Streptomycineae; Streptomycetaceae; Streptomyces.
 NC NCBI_TaxID=38315;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ISP 5395;
 RA Hoshi K.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 XC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
 XC intermediary of phytofluene by the symmetrical introduction of two
 XC double bonds at the C-11 and C-11' positions of phytoene.
 XC -!- COFACTOR: FAD (Probable).
 XC -!- PATHWAY: Carotenoid biosynthesis.
 XC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 XC -----
 XC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D55723; BAA09537.1; -
 DR InterPro; IPR000759; Admrx_reductase.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac_phytoene_dh.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR008151; Phytin_dehydro.
 DR Pfam; PF01593; Amino_oxidase; 1.
 DR PRINTS; PR00419; ADMRTASE.
 DR PRODOM; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE DH; 1.
 KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 FT NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 508 AA; 54610 MW; 4BDEFC076D51CB5 CRC64;

Query Match 25.3%; Score 660.5; DB 1; Length 508;
 Best Local Similarity 34.7%; Pred. No. 7.2e-46;
 Matches 173; Conservative 69; Mismatches 213; Indels 43; Gaps 11;

QY 6 VIGAGGGLALALRLQAAGIPVLLLBQRDKPGRAVYVQEGFTFDAGTWTIDPSAIEE 65
 Db 13 VVGAGLAGLAAHLHLCAGRSVTVVEQGVPGGRAGLLETDGFRVDTGTVTWIDVED 72
 QY 66 LFALAGKQLKDYVELLPVTPFYELCWESKVFYNDQALQLEAQIQFNPDRVAGY--- 121
 Db 73 AFAVGERMADRLRLAPAYRAREADGSLDVTDTGAAMEAAVEQFAGARQAVGYRL 132
 QY 122 -----RAFLDYSAVFNEGYLKLGTVPF--LSPKMLRAAPQAKLQAWESVYS 168
 Db 133 RIMLERLYRVQMRRTD-----ANFDS-----PFQLVHFDLAR-----LAALGGFRELDA 177
 QY 169 KVAGYIIDEHLRQAFSPHLLVGNPNFATSSIIYTLHALERWGVNFPFGGTGALVNGMI 228
 Db 178 RIGHFVSDERLRRVFSQALYAGVPPARALAAVAVIYMDTVAGVYFPRGGMHALPRAMA 237
 QY 229 KLFDLGGEVVLNRVSHMETVGKQAVOLEGRRPETCAVASNADVHTTRDILLSQHP 288
 Db 238 DAAADAGASTRYGQSVTRFLRSRGRVTAV-VTDQRIADAVVLTDPDLPVSYR-LLGRSP 295
 QY 289 AAAQAKKLOSKMNSLFLVYFGLNHHDDLAHTVCFGRYRELHIEIFNHGDLAEF 348
 Db 296 -----HEPLPLRHSFSAVILHAGTDTWDLAHTTISFGAAWNTFHELTTRTGLMSDP 349
 QY 349 SLYLHAPCVTDPDLPAGCGSYVYVLAPEVH--LGTANLDWAVEGPRRLDRIDFYLEQHYN 406
 Db 350 SLLITRPTASDPSLAPGSKHLHYVLAPEVH--LGTANLDWAVEGPRRLDRIDFYLEQHYN 409
 QY 407 PGLRSQVTHRMFTPEPFRDELNAWQSSAFSVEPILTQSAWFRPHNRKHIDNLYLVGAG 466
 Db 410 PGLGSAIEVGLVTPDVTWTAQGH-AGTPFSVAHTFTQTPFRNLRVGTNVNAVLAGCG 468
 QY 467 THPGAGIPGVIGSAKATA 484
 Db 469 TTPGVGVTVTVLISGLKAA 486

RESULT 9
 CRTI NEUCR STANDARD; PRT; 595 AA.
 ID CRTI_NEUCR
 AC P21334;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase) (Albino-1
 DE protein).
 GN AL-1.
 OS Neurospora crassa.

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
NCBI_TaxID=5141;
[1]
SEQUENCE FROM N.A.
STRAIN=74-OR23-1A / FGSC 987;
MEDLINE=90377195; PubMed=2144609;
Schmidhauser T.J., Lauter P.-R., Russo V.E.A., Yanofsky C.;
"Cloning, sequence, and photoregulation of al-1, a carotenoid
biosynthetic gene of Neurospora crassa";
Mol. Cell. Biol. 10:5064-5070(1990).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- INDUCTION: By photoinduction.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M57465; AAA33555.1; -
PIR; A35919; A35919.
InterPro; IPR002937; Amino_oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR000205; NAD_BS.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
Transmembrane.
NP_BIND 11 44 FAD (ADP PART) (POTENTIAL).
TRANSMEM 578 595 POTENTIAL.
SEQUENCE 595 AA; 66367 MW; 0FF3DF07328BD784 CRC64;
Query Match 24.4%; Score 635; DB 1; Length 595;
Best Local Similarity 31.3%; Pred. No. 1e-43;
Matches 158; Conservative 97; Mismatches 226; Indels 24; Gaps 11;
6 VIGAGFGGLAIAIRLQAGIPVLLLEQDKPGGR-AVYOEQGFEDAGTPTVITDPSAIE 64
12 IVGAGAGGIAVAIRLAKAGVDVTVLEKNDFTGRCSLIHTKAGVRFQGSLLLLPOLFR 71
65 ELFALAGKQL-KDYVELLPVTPFRLCWESKVNVDNDQALEAQIQNPRDVAGYRA 123
72 ETFEDLGTLEQEDVELLQCPPTNINFSQKRFSTPTDNTATMKVLEKEWEGPD--GFR 129
124 FLDSRAVFNDSGLYKLGTVPLSPFLSKDMLRAA-POLA-----KLQAMRSVYSKVAGYIEDE 177
130 YLSKLAEGHQHYETSLRHVRHNFKSILELADPLVTLMLALHPFESIHWRAGRFTKD 189
178 HLKQAPSHSLVGNPPATSSIVTLIHALERKGVWPPRGGTGALVNGMKLFDQLGGE 237
190 RMQRVFTATVMGMSPDAPATYSLQYSLADGIWVPRGGFHKVLDALVKGIRMGVYK 249
238 VILNARVSHMTVG-----DKIQAVOLEGRRRETCAVASNADVDVHTYRDLASQHPAA 291
250 YRLNTGVSVQVTDGKGNGKPRATGVQNGEVNLVADLVVNDLVVYNNLLPKETGGI 309
292 KQ-AKKLASKMSNLSFLYFGLNHHHDQLAHTVCFGRPYRELIIHFHNDGLAEDFSL 350
310 KKYANKLNRRASCSSISFYSLSGMAKELTHNI FLAEEYKESFDAIFERQALPDPSF 369
351 YLHAPCVTDPSLAPBGCCSYVILAPVPHL-----GTANLDMAVEGRPLNDRIDYLEHYMP 407
370 YTHVPSVDPSAAPPDRDRAVYALVPVGHLLQNGQFELDWPTLVSKAPAGVLAITQAR--T 427

QY 408 GLR-SQVTHRMP-TFFDRDELNAWQSASFVPIITQSAWFRPHNRDKHIDMLYVGA 465
DB 428 GLSLSPILTEIBIVNTPTWETKFNLSKAILGLAHDFFNVLAFRPTKAQGMNAYFVGA 487
QY 466 GTHPCAGIPGVIGSAKATAGLMLSD 490
DB 498 STHPGTGPIVLAGAKITAEQILEE 512

RESULT 10

CRTI_MYXXA
ID CRTI_MYXXA STANDARD; PRT; 529 AA.
AC Q02861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) [Phytoene desaturase].
GN CARC.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cytophactariae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=MR403;
RX MEDLINE=93223667; PubMed=8467787;
RA Pontes M., Ruiz-Vazquez R.M., Murillo F.J.;
RT "Growth phase dependence of the activation of a bacterial gene for
carotenoid synthesis by blue light";
RL EMBO J. 12:1265-1275(1993).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
intermediaries of phytofluene, zeta-carotene and neurosporene by
the introduction of four double bonds (By similarity).
CC -!- COFACTOR: FAD (Probable).
CC -!- INDUCTION: By blue light.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M94727; AAA25390.1; -
PIR; S35306; S35306.
InterPro; IPR000759; Adnrx_reductase.
InterPro; IPR002937; Amino_oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR000205; NAD_BS.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
PRINTS; PR00419; ADXEDTASE
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
FT SEQUENCE 529 AA; 58420 MW; 53536A8DFD0D24BC CRC64;
Query Match 23.6%; Score 616; DB 1; Length 529;
Best Local Similarity 30.2%; Pred. No. 3.1e-42;
Matches 153; Conservative 99; Mismatches 218; Indels 36; Gaps 9;

QY 6 VIGAGFGGLAIAIRLQAGIPVLLLEQDKPGGRAYVYQGGFTF-----DAGTPTVI 57
DB 13 VVGAGPGGLSAAINLAGQGPRVTTVVXKDAVEGGRM-----KGLTLGASGEVAVDTGPSIL 67
QY 58 TDPNSIELPALAGKQLKDYVELLPVTPFRLCWESKVNVDNDQALEAQIQNPRD 117
DB 68 QLPGLVLEQIFRAARRRDEYVKKLPLDVNTKRVHFDGTHLDTTHLDRMEALAKFGRQ 127

```

QY 118 VAGYRAFLDYSAVENBGLYKL--GTVPFILSPKDMLEBAAPOLAKLOAWRSVYKVAQYIE 175
DB 128 ASALRQWVEDGREKYIAYQKICTSADNIGYAPWELAPTL-RFPWQTLXQLDGFTH 186
QY 176 DEHLRQAFSFLSLVGGNPFATSSIIYTLIHAREWGVWFPFGGTGALVNGMIKLPQDLG 235
DB 187 DDRVTYALAYSPLYGLHPTCCSVFSPFELAFGVWVHVEGFRSLSRGMKRCARDLG 246
QY 236 GEVVLNARVSHMETVGDKIQAQVLEDGRRETCAVASNADVVHYTYRDLLSQHAAKQAK 295
DB 247 ATRFMGTPEVKRVVDAGRAVGVKLVGGVEVDADAVVYNADLAVAAASLSI---PARAREGS 303
QY 296 K-----LQSKMSNSFLVLYEGLNHHHDQLAHTVCFGRPYRLIHIFNHDGL----- 344
DB 304 RLTDAAALERAKYSCSTFWAYGLDVTYADLPPLHLYLSESART----DRDALEDRHVD 358
QY 345 AEDSLYHARCVDTPSLAPGCGSYVYVAPVPHGLGTANLDWAVEGPRLEDRIFDYLEQH 404
DB 359 LEDPPFYVCNPGVTPDPSGAPAGHSTLYLVPTNTGRP--VDWVKTQEARERIPAMLEKV 417
QY 405 YMPGLRSOLVTHRMFTPFDFRDELNAWOGSAPFVEPILTQSAFRPHNRDKHIDNLYLVG 464
DB 418 GLKGVREHIREERTFAETHRDDFNVFRAVENLSHTWLQGLPLRPVKVKNDEIGLYFVG 477
QY 465 AGTHPGAGIPGVIGSAKATAGIMLED 490
DB 478 GGTHTPGSGLLTIMESANIAADYLRE 503

RESULT 11
CRTL_PHYBL STANDARD; PRT; 583 AA.
AC P54982;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
EN CARB.
OS Phycomyces blakesleeanus.
DC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
DC Phycomyces.
DX NCBI_TaxID=4837;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1555;
RA Ruiz-Hidalgo M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78434; CAA55197.1; -.
DR PIR; S43139; S43139.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR000205; NAD_B5.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytin dehydro; 1.
DR PROSITE; PS00982; PHYTOENE_DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;

```

```

KW Transmembrane.
FT NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
FT TRANSMEM 531 551 POTENTIAL.
SQ SEQUENCE 583 AA; 65983 MW; 808F682B12FB591 CRC64;

Query Match 22.9%; Score 595.5; DB 1; Length 583;
Best Local Similarity 29.1%; Pred. No. 1.6e-40;
Matches 146; Conservative 103; Mismatches 225; Indels 27; Gaps 12;

QY 2 KPTTVIGAGFGSLALAIQLQAAGIPVLLLEQRDKPGGRVAVYQEQGTFDAGPTVITDPS 61
DB 5 KEVVIIGAGAGTATARLARLEGKIVTVKKNPCCGRCSLINHNGHFRPQGSLSLYLMPK 64
QY 62 AIEELFALAGKQDYVELLPVTFYFRLCHWESGKVFNVDQAOLEAQIQOOFNPRDVAGY 121
DB 65 LFEEAFALDEKIDHVELLRCHNNYKVFHDDGDKIQIQLSSDLSRMKPEMERIEGPD--GF 122
QY 122 RAFLDYSPA--VFNEGYLKLGTVPFLSPKDM-----LRAAPQLAKLQAWRSVYSKVAGYI 174
DB 123 LRFLDFMKESHTHYEGGVEMAIKQ--NPFETIWLKIRLQYVPAFLRLHIFDFVYSRAAKYF 180
QY 175 EDEHLRQAFSFLSLVGGNPFATSSIIYTLIHAREWGVWFPFGGTGALVNGMIKLPQDL 233
DB 181 KTKKRWMAFTQSMYMGMSPYDSPAVYNELQYTFEASGIWYPKGFTNTVIQKLENIATEK 240
QY 234 LGGEVVLNARVSHMET--VGDKIQAQVLEDGRRETCAVASNADVVHYTYRDLLSQHAAA 291
DB 241 FGARFIYEAPVAKINTDDKGGKVTGVTLQSGEVLEADAVVCNADLVYAHNLL---PPCR 297
QY 292 KQAKCLQSKRMSNSLFVLYFGLNHHHDQLAHTVCFGRPYRLIHIFNHDGLAEDPSLY 351
DB 298 WTYNTLAELKLTSSISFYNSLKEVPSLDVHNLFLAEAFKESFDEIFTDHKMPSELSFY 357
QY 352 LHAPCVTDPSLAPGCGSYVYVAPVPHGLGTAN---LDWAVEGPRLEDRIFDYLEQHMPG 408
DB 358 VNLPSRIDPTAAPGKDSMIVLVPIGHMKSKTNEADYTMIVKEARQVLEVLRL--G 415
QY 409 LRS--QLVTH-RMTTPDFRDELNAWOGSAPFVEPILTQSAFRPHNRDK--HIDNLYLV 463
DB 416 LTNFIDLVEHEVNDPSINQKFKLWRGSLGLSHDVQLVLMFRPSTQDSTGYKMLPFV 475
QY 464 GAGTHPGAGIPGVIGSAKATA 484
DB 476 GASTHTPGTGVPIVLAGSKLTS 496

RESULT 12
CRTL_CERN
ID CRTL_CERN STANDARD; PRT; 621 AA.
AC P48537;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN PDH1.
OS Cercospora nicotianae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cercospora.
OX NCBI_TaxID=29003;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 18366;
RC MEDLINE=94368091; PubMed=8083820;
RX Ehrenschaft M., Daub M.E.;
RX "Isolation, sequence, and characterization of the Cercospora
RX nicotianae phytoene dehydrogenase gene."
RX Appl. Environ. Microbiol. 60:2768-2771(1994).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.

```


This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
EMBL; Y00142; CAA68336.1; -
PIR; S09189; S09189.
InterPro; IPR000759; Adrndx_reductase.
InterPro; IPR000205; NAD_BS.
InterPro; IPR008152; Phyt_n_dehydro.
PRINTS; PR00419; ADXNDTASE.
ProDom; PD139017; Phyt_n_dehydro; 1.
SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;
```

```
Query Match          6.6%; Score 171.5; DB 1; Length 469;
Best Local Similarity 23.3%; Pred No. 3.1e-06;
Matches 126; Conservative 64; Mismatches 212; Indels 139; Gaps 29;

1 MKPTTVICAGFGGLALAIRLOAGIPVLLLEQRDKPGRAYVYQ--EQGFED----AGP 54
1 MLDVAVVGAGPGLTAAVELARRGPPVAVPEAOCTVGGARTETLTLPGRHDPGSAHP 60

55 TVITDPSAIEELFALAGKQKDY-VEL-----LPVT-PFYRLCWESGRVFNVDNDQOLEA 108
61 LGINSP-----APRGPLERYGLEWHLPGLPMAHPF-----PDGSAAVLSR 101

109 QIQO-----FNPRDVAGYRAFLDYGRAVFNEGYKLGTVPLSFKDM-LRAAPQ----LAK 159
102 SVGETAASFQADAGPYRRLI-----ERFLPRWDTLARDFMSLFLTLALPRDPVTILAR 153

160 L-----QAWRSVYSKVAGVIEDEHLRQAPS-----FHSLL-VGQNPATSSIVTLIHAL- 207
154 FGLVGLPEPTW-----LMRRFRDEKAKTLFAGLVAVHMAPLGG--FATGAI-GLVPALA 204

208 --EREWGVWFRGGTGAIVNGMIKLFQDLGCEVVVLNARVSHMETVGDKIQAVOLEDGRRF 265
205 AHARGWV--ARGSQSISDALTYLKDLOGAVHTDYEVKGLD-----DLPPARAY 253

266 ETCVASNADVVHTVRDILLSQHPAAAKQAKLQSK-----RMSNSLFVLYFCLNHHHDQLA 321
254 -----VLDTSPTALARIAGLGSYVANYRYGSPVKIDYALDGPVFWTA 296

322 HH-----TVCFGPYREL---IHEIFNHDGLAEDFSIYLHAPCVTDPSLAPEGGSYYVL 373
297 EEPKSAGTVQIGADSTEIGAAHAPSGTDRAPERPFITVQPSVADPTRAPAGKHVFWAY 356

374 APVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVLSQVTHRMFTPTDFRDELNAWQG 433
357 CHVENGWGDLTDAME--QLER-----FAPGFRDRLARATAGPAELAARNANYVG 406

434 SAFP-----VEPILTOSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIG--SAK 481
407 GDISGAVSGQLLLRPKISLFPYSTPH-----PAVFICSSATPPGPGVHGMSGHNAAK 460

482 A 482
461 A 461
```

arch completed: February 29, 2004, 14:45:12
b time : 10.2921 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 13.6946 Seconds
(without alignments)
3455.835 Million cell updates/sec

file: US-09-941-947a-32

fect score: 2606
quence: 1 MKPTTVIGAGFGGLAIRL.....IPGVIGSAKATAGLMLDLI 492

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tail number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase :
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	2453	94.1	492	2	S52586	phytoene dehydroge
2	2438	93.6	492	2	D37802	phytoene dehydroge
3	2023	77.6	492	2	A33120	phytoene dehydroge
4	1099.5	42.2	511	2	T50910	phytoene dehydroge
5	1051	40.3	548	2	C75466	phytoene dehydroge
6	1007.5	38.7	524	2	A32617	phytoene dehydroge
7	959.5	36.8	518	2	S49620	phytoene dehydroge
8	932.5	36.6	518	2	T50745	phytoene dehydroge
9	757.5	29.1	517	2	S32169	hypothetical prote
10	752	28.9	517	2	T31463	probable diaphophyt
11	635	24.4	595	2	A35919	carotenoid biosynt
12	629.5	24.2	523	2	T36968	probable phytoene
13	624	23.9	502	2	B90061	squalene synthase
14	616	23.6	529	2	S35306	phytoene dehydroge
15	604	23.2	499	2	AG2209	zeta-carotene desa
16	595.5	22.9	583	2	S43139	phytoene dehydroge
17	592	22.7	499	2	A43324	zeta-carotene desa
18	567	21.8	490	2	AH1199	phytoene dehydroge
19	562	21.6	582	2	T46822	phytoene desaturas
20	549.5	21.1	514	2	B69108	phytoene dehydroge
21	541.5	20.8	512	2	H84320	phytoene dehydroge
22	530	20.3	621	2	T48646	Phytoene dehydroge
23	493	18.9	448	2	B55348	crtn protein - Sta
24	489	18.8	488	2	AF1557	phytoene dehydroge
25	482	18.5	497	2	E90061	hypothetical prote
26	466	17.9	528	2	T51119	phytoene desaturas
27	433.5	16.6	536	2	B84327	Phytoene dehydroge
28	427.5	16.4	486	2	S49624	methoxyneurosporen
29	423.5	16.3	495	2	S23333	methoxyneurosporen

30	411	15.8	485	2	S32171	hydroxyneurosporen
31	388.5	14.9	486	2	T50749	methoxyneurosporen
32	380.5	14.6	542	2	S76617	hypothetical prote
33	375	14.4	565	2	A12273	hypothetical prote
34	369.5	14.2	494	2	S04406	methoxyneurosporen
35	343	13.2	454	2	A99470	phytoene dehydroge
36	342.5	13.1	498	2	H83880	hypothetical prote
37	322	12.4	506	2	AB2064	hypothetical prote
38	316	12.1	511	2	E75561	probable phytoene
39	313	12.0	525	2	T50893	methoxyneurosporen
40	311	11.9	501	2	S75951	hypothetical prote
41	292.5	11.2	587	2	AB6203	hypothetical prote
42	289	11.1	543	2	G87635	phytoene dehydroge
43	276.5	10.6	518	2	G90413	phytoene dehydroge
44	274.5	10.5	506	2	T34971	probable carotenoi
45	259	9.9	380	2	A12185	hypothetical prote

ALIGNMENTS

RESULT 1

S52586

phytoene dehydrogenase (EC 1.3.-.-) - Erwinia herbicola

C:Species: Erwinia herbicola

C>Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999

C:Accession: S52586

R:Lin, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tni000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52586

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-492 <Lin>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21263.1; PID:G148397

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: phytoene dehydrogenase

C:Keywords: oxidoreductase

Query Match 94.1%; Score 2453; DB 2; Length 492;
Best Local Similarity 93.7%; Pred. No. 5.9e-189;
Matches 461; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

QY	1	MKPTTVIGAGFGGLAIRLQAGIPVLLLEQDKPGRAYVYCEGFTDAGTWTDP	60
DB	1	MKPTTVIGAGFGGLAIRLQAGIPVLLLEQDKPGRAYVYCEGFTDAGTWTDP	60
QY	61	SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFNYDNDQAQLEAQIQQFNPRDVAG	120
DB	61	SAIEELFTLAGKQKDYVELLPVAPFYRLCWESGKVFNYDNDQAQLEAQIQQFNPRDVEG	120
QY	121	YRAFIDYSRAVFNHGYKLGTVFPLFKDMLRAAPOLAKLOAWRSYVKVAGYIEDEHLR	180
DB	121	YRQFLDYSRAVFNHGYKLGTVFPLFKDMLRAAPOLAKLOAWRTYSKVASYIEDEHLR	180
QY	181	QAFSFHLLVGGNPPFATSSITLIHALERENGWVFPFGGTGALVNGMKLFDGLGGEVVL	240
DB	181	QAFSFHLLVGGNPPFATSSITLIHALERENGWVFPFGGTGALVNGMKLFDGLGGEVVL	240
QY	241	NARVSHMETVGDKIQAQVLEDGRRPFTCAVASNADVVHYTRDLLSOHPAAAKQKQSK	300
DB	241	NARVSHMETVGDTEAVHLEDGRRPFTCAVASNADVVHYTRDLLSOHPAAVQSKQKQSK	300
QY	301	RMSNSLPVLYFGLNHHDDLAHTVCGFPYRLIHEIFNHDLGLAEDFSLYLHAPCVTPD	360
DB	301	RMSNSLPVLYFGLNHHDDLAHTVCGFPYRLIHEIFNHDLGLAEDFSLYLHAPCVTDS	360
QY	361	SLAPEGCGSYVILAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVITRMPT	420
DB	361	SLAPEGCGSYVILAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVITRMPT	420
QY	421	PFDFRDLNANQSSASVEPILTQSAWFRPHNEDKIDNLYLVGAGTHPCAGIPGVIGSA	480

```
b 421 PFDFRQLNAYQSSAFSVEPVLQTQSAWFRPHNRDKTINNLVLVGAGTHPGAGIPGVIGSA 480
y 481 KATAGLMLEDLI 492
b 481 KATAGLMLEDLI 492

RESULT 2
37802
hyoene dehydrogenase (EC 1.3.-.-) crti - Erwinia uredovora
;Species: Erwinia uredovora
;Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
;Accession: D37802
;Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashige,
;Bacteriol. 172, 6704-6712, 1990
;Title: Clucidation of the Erwinia uredovora carotenoid biosynthetic pathway by function
;Reference number: A37802; MUID:91072214; PMID:2254247
;Accession: D37802
;Status: preliminary
;Molecule type: DNA
;Residues: 1-492 <MIS>
;Cross-references: GB:D90087; NID:G216681; PIDN:BAAL127.1; PID:G216685
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match 93.6%; Score 2438; DB 2; Length 492;
Best Local Similarity 93.3%; Pred. No. 9.5e-188;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

y 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
b 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
y 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
b 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
y 121 YRAFLDYSRAVFNQVYKLGTVFPLSFKDMRAAPQAKLQAWRSVYKVGAYVIEDEHLR 180
b 121 YRAFLDYSRAVFNQVYKLGTVFPLSFKDMRAAPQAKLQAWRSVYKVGAYVIEDEHLR 180
y 181 QAFSPHLLVGGNPFATSSITLILHALEREWGVPFGGCTGALVNMKILFQDLGGEVVL 240
b 181 QAFSPHLLVGGNPFATSSITLILHALEREWGVPFGGCTGALVNMKILFQDLGGEVVL 240
y 241 NARVSMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQAAPAAKQKLOSK 300
b 241 NARVSMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQAAPAAKQKLOSK 300
y 301 RMNSLFLVYFGLNHHDDQLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
b 301 RMNSLFLVYFGLNHHDDQLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
y 361 SLAPEGCGSYVILAPVPHLGTANLDAWVGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
b 361 SLAPEGCGSYVILAPVPHLGTANLDAWVGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
y 421 PFDFRDLNANQGSASFSEVPILTQSAWFRPHNRDKHIDNLVLVGAGTHPGAGIPGVIGSA 480
b 421 PFDFRDLNANQGSASFSEVPILTQSAWFRPHNRDKHIDNLVLVGAGTHPGAGIPGVIGSA 480
y 481 KATAGLMLEDLI 492
b 481 KATAGLMLEDLI 492

RESULT 3
33120
hyoene dehydrogenase (EC 1.3.-.-) - Erwinia herbicola
;Alternate names: phytoene desaturase
;Species: Erwinia herbicola
;Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 29-Sep-1999
```

```
C;Accession: A39273; A33120
R;Armstrong, G.A.; Alberti, M.; Heatet, J.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990
A;Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in non
A;Reference number: A39273; MUID:91088634; PMID:2263648
A;Accession: A39273
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-492 <ARM>
A;Cross-references: GB:M38423; NID:G148401; PIDN:AAA24820.1; PID:G148402
A;Note: the authors translated the codon CAG for residue 181 as Phe, TCC for residue 18'
A;Note: Strain Bho10; ATCC 39368
C;Genetics:
C;Superfamily: phytoene dehydrogenase
C;Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 77.6%; Score 2023; DB 2; Length 492;
Best Local Similarity 76.8%; Pred. No. 2e-154;
Matches 377; Conservative 42; Mismatches 72; Indels 0; Gaps 0;

Qy 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
Db 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
Qy 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
Db 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
Qy 121 YRAFLDYSRAVFNQVYKLGTVFPLSFKDMRAAPQAKLQAWRSVYKVGAYVIEDEHLR 180
Db 121 YRAFLDYSRAVFNQVYKLGTVFPLSFKDMRAAPQAKLQAWRSVYKVGAYVIEDEHLR 180
Qy 181 QAFSPHLLVGGNPFATSSITLILHALEREWGVPFGGCTGALVNMKILFQDLGGEVVL 240
Db 181 QAFSPHLLVGGNPFATSSITLILHALEREWGVPFGGCTGALVNMKILFQDLGGEVVL 240
Qy 241 NARVSMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQAAPAAKQKLOSK 300
Db 241 NARVSMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQAAPAAKQKLOSK 300
Qy 301 RMNSLFLVYFGLNHHDDQLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
Db 301 RMNSLFLVYFGLNHHDDQLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
Qy 361 SLAPEGCGSYVILAPVPHLGTANLDAWVGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
Db 361 SLAPEGCGSYVILAPVPHLGTANLDAWVGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
Qy 421 PFDFRDLNANQGSASFSEVPILTQSAWFRPHNRDKHIDNLVLVGAGTHPGAGIPGVIGSA 480
Db 421 PFDFRDLNANQGSASFSEVPILTQSAWFRPHNRDKHIDNLVLVGAGTHPGAGIPGVIGSA 480
Qy 481 KATAGLMLEDLI 491
Db 481 KATAGLMLEDLI 491

RESULT 4
T50910
phytoene dehydrogenase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C;Accession: T50910
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynt
A;Reference number: Z25270
A;Accession: T50910
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-511 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BAAG4063.1
```


Experimental source: strain IL144
Genetics:
Gene: crtII
Superfamily: phytoene dehydrogenase

Query Match 42.2%; Score 1099.5; DB 2; Length 511; Gaps 1;
Best Local Similarity 42.3%; Pred. No. 3e-80;
Matches 204; Conservative 100; Mismatches 177; Indels 1; Gaps 1;
6 VTGPGGLALAIKLAQAGIPVLLLEQRDPKGGRAVYVQ-QGFTTQAGTWTIDP 60
17 KTAIVIGAGFGGLAIGIRLOSGLGDTTILRLRDPGPGRAYOKRTPDGIVDMGTIVTVP 76
61 SAIELFA-----LAGKQKLD-----YVELLPVTPFYRLCWES 93
77 HFIEELFALERDRAGLDAPDYPPEVLSSGERVKGSGPHTSRVYTLVLPILFYRIVFDH 136
94 GKVFYNDQAOLEAQICQPNPRDVAGYRAFLDYSRVFNVEGVYKLGTVFLSPKDMIRA 153
137 GTYFDYDGDPESTRQIAELAPGLAGYERFHADAFAIRRGFLGELGYTHFGDVPTMLRV 196
154 APOLAKIQAWRSVYSKVAGYIIEDEHLRQAFSFHLLVGGNPPATFSIYTLIHALLERWGV 213
197 VPDLLKDAVTLFSTFSKYFQSKLRQVFSFETLLVGGNPLSVPAIYAMHFVEKWTGI 256
214 WPPGGTGALVNGMIKLFQDILGGEVVLNARVSHMETVGD-----KIQAVOLEDGRPETC 268
257 HYAMGGTGALVGVQKFEELGGAIYRGAGVDEVLDGNDLPGRKTARGVLESGEELRAD 316
269 AVASNADVHTYRDLISQHPAAAKQAQKLOSKMSN--SLFVLYFGLNHHHD-QLAHTTV 325
317 LVASNGDWANTY--LKRVRPSARLVNSDLAVKAASSMSLLVYFGRGDDLPKXHHNI 374
326 CFGPRYRELIIHIFNHDGLADPSLYLHAPCVTDPSLAPEGCGSYVYVLAHPHGLGTANLD 385
375 LLGPRYEAALLSEIPGTRKLGEDFSQYLRVPTLTDPALAPAGHAAAYTLVPVPHNG-SGID 433
386 WAVEGPRRLRRIIDYLEQH-YMPCLRSQLVTHRMFTFDPDRDELNAWQGSFAFSVEPILTO 444
434 MDVGGPKLAELAAADIERRGLIPGLRRLTHFFITPDYFAGTLDLSYLGNAFGPEPRLVQ 493
445 SAWFRPHNRKHIIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLED 490
494 SAFFRPHNRSEDHNFYLVGAGAGPGAGTSPVMSVAQMTARLIAED 539
RESULT 6
A32617
phytoene dehydrogenase (EC 1.3.-.-) - Rhodobacter capsulatus
N:Alternate names: phytoene desaturase
C:Species: Rhodobacter capsulatus
C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
C:Accession: A32617; S04402
R:Bartley, G.E.; Scolnik, P.A.
J. Biol. Chem. 264, 13109-13113, 1989
A:Title: Carotenoid biosynthesis in photosynthetic bacteria. Genetic characterization of
A:Reference number: A32617; MUID:89327279; PMID:2546948
A:Accession: A32617
A:Molecule type: DNA
A:Residues: 1-524 <BAR>
A:Cross-references: GB:J04969; NID:G340554; PIDN:AAAS0313.1; PID:G556397
R:Armstrong, G.A.; Alberti, M.; Leach, F.; Hearst, J.E.
Mol. Gen. Genet. 216, 254-268, 1989
A:Title: Nucleotide sequence, organization, and nature of the protein products of the ca:
A:Reference number: S04401; MUID:89313663; PMID:2747617
A:Accession: S04402
A:Molecule type: DNA
A:Residues: 1-524 <ARM>
A:Cross-references: EMBL:X52291; NID:G45996; PIDN:CAA36533.1; PID:G45998
A:Note: translation of codons 1-33 is not given
A:Note: the authors translated the codon GTG for residue 34 as Met
C:Genetics:
C:Gene: crtII
C:Superfamily: phytoene dehydrogenase
C:Keywords: carotenoid biosynthesis; oxidoreductase
Query Match 38.7%; Score 1007.5; DB 2; Length 524;
Best Local Similarity 41.5%; Pred. No. 7.6e-73;
Matches 204; Conservative 89; Mismatches 194; Indels 5; Gaps 3;
1 MKPTTVICAGFGGLAIIKLAQAGIPVLLLEQRDPKGGRAVYVQ-QGFTTQAGTWTIDP 60

SULT 5

5466
ytene dehydrogenase - Deinococcus radiodurans (strain R1)
Species: Deinococcus radiodurans
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
Accession: C75466
White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
; Smith, H.O.; Venter, J.C.; Fraser, C.M.
ience 286, 1571-1577, 1999
Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
Reference number: A75250; MUID:20036896; PMID:10567266
Accession: C75466
Status: preliminary
Molecule type: DNA
Residues: 1-548 <WHI>
Cross-references: GB:AF001940; NID:G6458577; PIDN:AAF10439.1; PID:G645858
Experimental source: strain R1
Genetics:
Gene: DR0861
Map position: 1
Superfamily: phytoene dehydrogenase
Query Match 40.3%; Score 1051; DB 2; Length 548;
Best Local Similarity 43.5%; Pred. No. 2.6e-76;
Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9;


```

b 8 MGRAVVIAGLGLAANRLGANGKYVTVVDRDLRDPGGRGSSITKGGHRFDLGPITVTP 67
y 61 SAIEHELALAGKQLKDYVELLPVTPYRLCWESGKVFNVDNDQAOLEAQIQFNPDRVAG 120
b 68 DRLEBLWADCGRDFDKVSLVPMPEYITDFDGEKYTAIGDADAKYAEVARIISPGDVEG 127
y 121 YRAFLDYSAVFNVEGKLGTVPLFSFKDLRAAPQAKLQAWRSVYKVGAGYIEDEHLR 180
b 128 FRHFMDAKARIEFGYENLGRKPMKLDLKVLPFGWLRADRSVGHAKGVKDDHLR 187
y 181 QAFSFLSLVGNPPATSSITYLIIHALEREWGVPFRGGTGALVNGMIKLFODLGEVVL 240
b 188 FALSFPFLFEGGDPFHVTSYILVSQLEKFGVHAIGVQAIDAMAKVITDQGENRL 247
y 241 NARVSHMETVGD-KIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAKQAKLQS 299
b 248 NTEVDILVSRDGKATGIRLMOTELPAQVVVSNADAGHTYKLLRNDRDWTDEKLDK 307
y 300 KRWNSLFLVLYPGLN--HHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPC 356
b 308 KRWMSGLFVYFGTKTAKXMKDVGHHTVVVGVPRYKEHVQDIFIKGELAEDMSLYVHRPS 367
y 357 VTDPSLAPEGCGSYVYLAAPVPHLGTAN-LDWAIVGPRLRDRIDFDYLEQHYMPGLSOLVT 415
b 368 VTDPTAAPGDDITFYVLSPVNLDGPDGVDWSVEABKYAKVKEERLEFGVAEKITE 427
y 416 HRMFTPFDFDELNAMQGSFAFSEPIILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPG 475
b 428 EVVTFPTFDRYLSPLGAGFSLEPRILQSAWFRPHNASEEVDGLYLVGAGTHPGAGVPS 487
y 476 VIGSAKATAGLM 487
b 488 VIGSGSLVAQMI 499

RESULT 7
49620
hycoene dehydrogenase (EC 1.3.-.-) - Rhodobacter sphaeroides
;Alternate names: phytoene desaturase
;Species: Rhodobacter sphaeroides
;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 13-Sep-1998
;Accession: S49620
;Author: K. P. Cogdell, R. J. Takaiuchi, S. J. Hunter, C. N.
submitted to the EMBL Data Library, November 1994
;Description: The complete DNA sequence, specific TMS insertion map and gene assignment
;Reference number: S49619
;Accession: S49620
;Status: preliminary
;Molecule type: DNA
;Residues: 1-518 <LAN>
;Cross-references: EMBL:X82458; NID:G575405; PID:G575407
;Genetics:
;Gene: ctri
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match 36.88; Score 959.5; DB 2; Length 518;
Best Local Similarity 40.98; Pred. No. 5.3e-69;
Matches 199; Conservative 79; Mismatches 201; Indels 7; Gaps 5;

y 6 VIGAGFGLALAIRLQAGIPVILLLEQRDKPGGRAYVYQEGFTFDAGTIVTIDPSAIBE 65
b 15 VIGSGLGLAAANRLGAKGWRVTVIDKLDVPGGRGSSITQEGHRFDLGPITVTPQSLRD 74
y 66 LPALAGKQLKDYVELLPVTPYRLCWESGKVFNVDNDQAOLEAQIQFNPDRVAGYRAPL 125
b 75 LMTCCGRDFDADVELKPIDPFYEVWRPDSGSHFTVROSTAMKAEVARLSPGVDAGTEKPL 134
y 126 DYSAVFNVEGKLGTVPLFSFKDLRAAPQAKLQAWRSVYKVGAGYIEDEHLQAFSP 185
b 135 KDSKRYWFGVEDLGRSMHKLWDLIKVLPFGWLRADRSVGHAKGVKDELRWALS 194
y 186 HSLLVGNPFPATSSITYLIIHALEREWGVPFRGGTGALVNGMIKLFODLGEVVLNARS 245
b 188 FALSFPFLFEGGDPFHVTSYILVSQLEKFGVHAIGVQAIDAMAKVITDQGENRL 247
y 241 NARVSHMETVGD-KIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAKQAKLQS 299
b 248 NTEVDILVSRDGKATGIRLMOTELPAQVVVSNADAGHTYKLLRNDRDWTDEKLDK 307
y 300 KRWNSLFLVLYPGLN--HHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPC 356
b 308 KRWMSGLFVYFGTKTAKXMKDVGHHTVVVGVPRYKEHVQDIFIKGELAEDMSLYVHRPS 367
y 357 VTDPSLAPEGCGSYVYLAAPVPHLGTAN-LDWAIVGPRLRDRIDFDYLEQHYMPGLSOLVT 415
b 368 VTDPTAAPGDDITFYVLSPVNLDGPDGVDWSVEABKYAKVKEERLEFGVAEKITE 427
y 416 HRMFTPFDFDELNAMQGSFAFSEPIILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPG 475
b 428 EVVTFPTFDRYLSPLGAGFSLEPRILQSAWFRPHNASEEVDGLYLVGAGTHPGAGVPS 487
y 476 VIGSAKATAGLM 487
b 488 VIGSGSLVAQMI 499

```

```

Db 195 HPLFIGGDPFNVTSMYILVSQLEKFGVHAIGVQAIAAAMAKVIEGGGSRFMMTEVD 254
Qy 246 HMETVGDKIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAKQAKLQSKEMSN 305
Db 255 EILVEKGTATGVLASGEVLRAGLVVSNADAGHTYKLLRNDRDWTDEKLDK 314
Qy 306 LFVLYFGLNHHH---DQLAHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPSL 362
Db 315 LFVWYFGTKGTGMMPDVGHHTVINAAPRYKGLVEDIFLKGKLAOKMSLYIHRPSITDPTV 374
Qy 363 APGCGSYYVLAAPVPHLGTAN-LDWAIVGPRLRDRIDFDYLEQHYMPGLSOLVTHRMPTP 421
Db 375 APGDDTTFALSPVPHLQKQAPVDVMAVAPRESVLEVEQS-MPGIGERIGPSLVFPT 433
Qy 422 FDRDE-LNAAQGSFAFSEPIILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPG 480
Db 434 ETRDRYLSW-GAGFSIEPRILQSAWFRPHNISEVANLFLVAGTHPGAGVPGVIGSA 492
Qy 481 KATAGL 486
Db 493 EVNAKL 498

RESULT 8
T50745
phytoene dehydrogenase (EC 1.3.-.-) [imported] - Rhodobacter sphaeroides
;Species: Rhodobacter sphaeroides
;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
;Accession: T50745
;Author: R. Choudhary, M. J. Kaplan, S.
Nucleic Acids Res. 28, 862-867, 2000
;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides
;Reference number: 225222; MUID:20115911; PMID:10648776
;Accession: T50745
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-518 <CHO>
;Cross-references: EMBL:AF195122; PIDN:AAF24289.1
;Experimental source: strain 2.4.1
;Genetics:
;Gene: ctri
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match 36.88; Score 952.5; DB 2; Length 518;
Best Local Similarity 40.98; Pred. No. 1.19e-68;
Matches 199; Conservative 78; Mismatches 202; Indels 7; Gaps 5;

Qy 6 VIGAGFGLALAIRLQAGIPVILLLEQRDKPGGRAYVYQEGFTFDAGTIVTIDPSAIBE 65
Db 15 VIGSGLGLAAANRLGAKGWRVTVIDKLDVPGGRGSSITQEGHRFDLGPITVTPQSLCD 74
Qy 66 LPALAGKQLKDYVELLPVTPYRLCWESGKVFNVDNDQAOLEAQIQFNPDRVAGYRAPL 125
Db 75 LMTCCGRDFDADVELKPIDPFYEVWRPDSGSHFTVROSTAMKAEVARLSPGVDAGTEKPL 134
Qy 126 DYSAVFNVEGKLGTVPLFSFKDLRAAPQAKLQAWRSVYKVGAGYIEDEHLQAFSP 185
Db 135 KDSKRYWFGVEDLGRSMHKLWDLIKVLPFGWLRADRSVGHAKGVKDELRWALS 194
Qy 186 HSLLVGNPFPATSSITYLIIHALEREWGVPFRGGTGALVNGMIKLFODLGEVVLNARS 245
Db 195 HPLFIGGDPFNVTSMYILVSQLEKFGVHAIGVQAIAAAMAKVIEGGGSRFMMTEVD 254
Qy 246 HMETVGDKIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAKQAKLQSKEMSN 305
Db 255 EILVEKGTATGVLASGEVLRAGLVVSNADAGHTYKLLRNDRDWTDEKLDK 314
Qy 306 LFVLYFGLNHHH---DQLAHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPSL 362
Db 315 LFVWYFGTKGTGMMPDVGHHTVINAAPRYKGLVEDIFLKGKLAOKMSLYIHRPSITDPTV 374

```



```
;Title: Cloning, sequence, and photoregulation of al-1, a carotenoid biosynthetic gene
;Accession: A35919
;Reference number: A35919; MUID:90377195; PMID:2144609
;Status: preliminary
;Molecule type: DNA
;Residues: 1-595 <SCH>
;Cross-references: GB:M57465, GB:M33867; NID:G168746; PIDN:AAA33555.1; PID:Q168747
;Note: the authors translated the codon CTC for residue 157 as Glu, and GAG for residue 158 as Asp.
;Superfamily: phytoene dehydrogenase
;Keywords: transmembrane protein
```

Query Match	24.4%;	Score 635;	DB 2;	Length 595;
Best Local Similarity	31.3%;	Prod. No. 7.4e-43;		
Matches 158;	Conservative	97;	Mismatches 226;	Indels 24;
Gaps 11;				
y	6	VIGAGGGALAIRQAAGIPVLLLEORDKPGGR-AYVYQSGGTFFDAGPTVITDPSAIE	64	
c	12	IVGAGGGIIVAAARLAKAGVDVTVLEKQDFTGGRCSLIHTKAGYRFQGPSLLLPGLFR	71	
y	65	ELFALAGKOL-KDYVELLPVTFYFALCWESKVFYNDQDLQAELAQIQFNPDRVAGYRA	123	
c	72	ETPFEDLTLEQEDVELLQCPENYNIWFSDGKRFSPPTDNATMKVIEKMGPD--GFRF	129	
y	124	FLDYSRAVNEGYKLGTVPFLSPFQMLRAA--POLA-----KLQAWRSVYSKVAGYIEDE	177	
c	130	YLSWLAEGHQHYETSLRHLVHNFHFKSILELADRLVLTLLMALHPFSSIWHRAGRYFKTD	189	
y	178	HLRQAFSESLVAGGNPPATSIYVTLIHALREWCVPFPRGCTGALVNGMLKLPDQLGGE	237	
c	190	RMQRVTFATMNGMSPDPAPATYSLLQYSELAEGIWTPRGGFTHKVDALVKIGERGVK	249	
y	238	VVLNARVSHMETVG-----DKIQAVQLEDGRRFETCAVSNADVHTYTRDLSQHPAAA	291	
c	250	YRLNTGVSQVLTGCKNGKKPKATGVQLENGEVLNADLVVYNADLVYTNLLPKEIGGI	309	
y	292	KQ-AKKLQSKRMNSLTVLYGLNHHHQLAHHTVCFGRYRELIIHEIFNHDGLAEFSL	350	
b	310	KXVANKLNKAKSCSSISFYWSLSGMAKELETHNIFLAEYKESFDALFERQALPDPPSF	369	
y	351	YLHAPCVTPESLAPEGCGSYVYLAVPVHL--GTANLDNAVEGRLRDRIDFYLEQHYMP	407	
b	370	YIHVPSRVPSAAPPDRDRAVIALVPVGHLLQKQPELDWPTLVSKARAGVLIATQAR--T	427	
y	408	GLR-SQLVTHRMF-TPFDRLDELNAWQSASFVEPILTQSAFWFPHNRKDHIDLVLVGA	465	
b	428	GLSLSPLIETEEIVNTPYTWETKFNLSKGAILLGLAHDFFNVLAFRPTKACQWDAFYVGA	487	
y	466	GTHPGAGIPGVIGSAKATAGLMLED	490	
y	488	STHPTGTGVPILVLAGAKITAEOLLEE	512	

RESULT 12

36968
robable phytoene dehydrogenase (phytoene desaturase) - Streptomyces coelicolor
;Species: Streptomyces coelicolor
;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
;Accession: T36968
;Submitted to the EMBL Data Library, September 1999
;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

```

;Reference number: Z21607
;Accession: T36968
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-523 <SEE>
;Cross-references: EMBL:AL08962; PIDN:CA853153.1; GSPDB:GN00070; SCOEDB:SCJ1.35
;Experimental source: strain A3(2)
;Genetics:
;Gene: crtG; SCOEDB:SCJ1.35
;Superfamily: phytoene dehydrogenase

Query Match      24.2%; Score 629.5; DB 2; Length 523;
Best Local Similarity 32.8%; Pred. No. 1.7e-42;

```

Matches	161;	Conservative	76;	Mismatches	235;	Indels	19;	Gaps
Qy	3	PT---VVIGAGFGGLAIAILQRAAGTPIVLLLEQRDKPCGGRAVYVQEQTFDAGTPTVTD	59					
Db	7	PTGHHVVVVGAGLSGALCALHLLGAGRRVTVVEDAGFGGSRGSRVGLGYELDTGPTVLTW	66					
Qy	60	PSAIEELPALAGKOLKDYVELLLPVTTPYRLCMESGKVFVFNVDNDQAOLEAQIOQF-NPRDV	118					
Db	67	PHLADAEFAAVGDSLRRRVELTALDPAYRACFADGSALDVHTDGEAMBAEVRFFAGPAQA	126					
Qy	119	AGYRAFILDYSRAVFNQYKLGTVPLFSFKDMLRAAPQAKLQA---WRSVYSKVAGYIE	175					
Db	127	AGYRDLRRWLRELYRAQWRFFITDNFDSPLQLH--PDLARLAALGGFGRLDGRIGRFLS	184					
Qy	176	DEHLROAFSPFHSLLVGCNPATSSIVYLIALEREMGVWFPFRGGTCALVNGMKILFQDLG	235					
Db	195	DELLRRVFSQALYAGVAPARALAAVAYIMDTVAGVWFPKGGMHALPRAMADAAATAG	244					
Qy	236	GEVLNARVSNHETVGDKIQAVOLEDGRRETCAVASNADVVHTVRDILLSQHPPAAKQAK	295					
Db	245	ADLRWSAEVKALERSAGRVRAVHLASGERIACDAVYLTCELSTAY-GLLGRAP-----R	297					
Qy	296	KIQKRMNSLFLVYGLNHHHDLQAHHTVCFGRYRRELIHEIFNHDGLAEDFSLYLHAP	355					
Db	298	RPARLRHSPSIVILNAGTDTWPHLAHTLSFGAAWERTFEELTRTGBELMSDPSLLITRP	357					
Qy	356	CVTDPSLAPEGCGSYVYLAVPH--LGTANLNDWAGEGPRLDRIRFDYLBQHTMPGLRSQL	413					
Db	358	ITTHDPALAPPGRHLHYVLAFCPTNDVGPDAAVWRDLGPYRBSLVGVLESRLGEGFADSV	417					
Qy	414	VTHRMFTFDFRDELNNAWQGSAPSVRPILTQSAWPRPHNRKHDINLNLVLVGAGTHPGAGI	473					
Db	418	QBELIVTPTLDWDAAQCHA-AGSPSVSHSTPAQTGFPRNLVRGLDNNVLAGCGTTFGVG	476					
Qy	474	PGVIGSAAKATA	484					
Db	477	PTVLVSGKLAA	487					

RESULT 13

B90061
squalene synthase [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Aug-2002
C:Accession: B90061
R:Kuroda M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaico, C.; Seki, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: B90061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-502 <KUR>
A:Cross-references: GB:BA000018; PID:gl3702511; PIDN:BA043652.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: crtN
C:Superfamily: phytoene dehydrogenase

Query Match	23.9%;	Score 624;	DB 2;	Length 502;
Best Local Similarity	31.7%;	Pred.No. 4.4e-421;		
Matches 159;	Conservative	22;	Mismatches 421;	Indels 24; Gaps 12

QY	6	VIGAFGGGLALAILRLQAAGIPVLLLEORDKEGRAYVYOEQFTTFDAGPTVITDSALIEE	65
Db	5	VIGAVTGLAAAARIASQGEHWTFEKNVNGVGRMNQLKKQDFTFDGTFITVAMDPVYKD	64
QY	66	LFALAGKQLKDYVILLPVTPTFYLRCWESGVFVYNDNDQALQAQIQCFNFRDVGAYRFL	125
Db	65	VFTAGKNVDYIEGRRLRYIYDVFPHDDRIYVPTDLAELQOMLESIEFGSTHGFMSPL	124

126 -----DYSRAVENEGVYKLGTVTFPLSEKDLRAAPOLAKIQAWRSVYKVGXIEDSH 178
125 TDVYKXIEAIRYFLERTYKPS-----DFYNM-TSVVQGAUKLTNLHADOLIEHYDNEK 179
179 LRQAFSPHSLVGGNPATTSIYTLIHALRENGVWPPRGGTGALVNGMIKLPQDLGGEV 238
180 IQKLLAFQTLVIGIDPKRPSLSYIIPMIEMFVGHFKGGYMGAAQLAKLNDLVNI 239
239 VLNARVSH-METVGDKIQAQVLE-DGRRFETCAVASNADVVHTYRDLLOH-PAKAKQA 294
240 ELNAIEQIITDPKPKADAKLVNGDIRKFD--KILCTADPPSVASSLMDPFAPIKKYPP 297
295 KKLOSXRMSNSFLVLYFGLN-HHHDQLAHTVCFGRYRELIIHFNHDLGLADFSLYLH 353
298 HKIADLDYSCAFLWYIGIDVDVTDQVRLHNVIFSDDFRGNIIEIF-EGRLSYDPSIYV 356
354 APCVTDSPLAPEGCGSVYVLAIPVPHLT-ANLDWAVEG--PRLADRFIDYLEO-HYMPGL 409
357 VPAVADKSLAPEGKGIYVLMPTPELKTGSDWSDBALTOQIKHIIYKRLATIEVFEDI 416
410 RSQLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
417 KSHIVSETIFPNDFQYTHAKFSAFGLAFTLAQSNYRPNQVSRDYKDLYPAGASTHP 476
470 GAGIPGVIGSKAKAFAGLWLEDL 491
477 GAGVPIVLTSAKITVDEMIXDI 498

SULT 14
5306
ytoene dehydrogenase (EC 1.3.-.-) - Myxococcus xanthus
Alternate names: phytoene desaturase
Species: Myxococcus xanthus
Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 29-Sep-1999
Accession: S35306; S27594
Fonfies, M.; Ruiz-Vazquez, R.; Murillo, P.J.
BO J. 12, 1265-1275, 1993
Title: Growth phase dependence of the activation of a bacterial gene for carotenoid sy
Reference number: S35306; MUID:93223667; PMID:8467787
Accession: S35306
Molecule type: DNA
Residues: 1-529 <FON>
Cross-references: EMBL:M94727; NID:g150079; PIDN:AAA25390.1; PID:g150080
Gene: carC
Superfamily: phytoene dehydrogenase
Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 23.68; Score 616; DB 2; Length 529;
Best Local Similarity 30.28; Pred. No. 2.1e-41;
Matches 153; Conservative 99; Mismatches 218; Indels 36; Gaps 9;
6 VIGAFGGLALALRQAAGIPVLLLEQDKPKGRAYVYQGGTF-----DAGFTVI 57
13 VVAGPGGLSAAINLACGPRVTVEKDAVGGRW-----KGLTLAGASGEAVDTGPSIL 67
58 TDPSAIELFALAGKQLKDVVELLPVTFPRLCWESKVFYNDNDQAQLEAQIQGNPRD 117
68 QLPGVLEQIFRAARLEDDYKLLPLDVNTRVHFWDGTHLDTTHHLDRAELAKEGPRQ 127
118 VAGYRAFIDYSRAVFNEGYKLL--GTVPFLSPKMLRAAPOLAKIQAWRSVYKVGXIE 175
128 ASALRQWEDCREKYGIAYQKFICTSDNGLGYAPWELAPTL-RFPKQWTLYRLQDGFPH 186
176 DEHURQAFSPHSLVGGNPATTSIYTLIHALRENGVWPPRGGTGALVNGMIKLPQDLG 235
187 DDRVYALAYPSKYLGLHPTTCSVSFVIFLEAFGVHVEGGFRELGRMWCARDIG 246
236 GEVVLNARVSHMETVGGKIQAQVLEDDGRRFETCAVASNADVVHTYRDLLOH-PAKAKQA 295
247 ATFRMGTFVKVRDAGAVGVKLVGGEVLDADVVVNADLAVALRSLLI---PAEREGS 303

QY 296 K-----LOSKEWNSLFLVLYGLNHHHDQLAHTVCFGRYRELIIHFNHDLG----- 344
DB 304 RUTDAALERAKYSCSTFMAYYGLDVTYADLPHHLYIYSESART-----DRDALEDRHVD 358
QY 345 AEDPSLYLHAFCVTDPSLAPEGCGSVYVLAIPVPHLGTANLDWAVEGFLRDRIFDYLEQH 404
DB 359 LEDPPYVNCNPGVTDPSGAPAGHSTLYLVPTNTGRP-VDWVTEQALREIRIPAMLEKV 417
QY 405 YMPGLARSOLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKHIDNLYLVG 464
DB 418 GLKGVREHIRERYFTATETWRDDFNVRGAVFNLSHTWLQGLPRLPKVKNORDIEGLYFVG 477
QY 465 AGTTPGAGIPGVIGSKAKATAGLMELED 490
DB 478 GGTTPGSGELLTIMESANTADYLTRE 503

RESULT 15
AG2509
zeta-carotene desaturase [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2509
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-499 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA078339.1; PID:g17135793; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7255
A:Genome: plasmid
C:Superfamily: phytoene dehydrogenase

Query Match 23.28; Score 604; DB 2; Length 499;
Best Local Similarity 29.8%; Pred. No. 1.8e-40;
Matches 153; Conservative 89; Mismatches 224; Indels 48; Gaps 9;
QY 2 KPTTVIGAGPGGLALALRQAAGIPVLLLEQDKPKGRAYVYQGGTFPAGTIVTDPS 61
DB 3 KKVAVIGAGPGGLATATRLAGLGTQVEIFEAERVGGRMERGFVDSYAFDTGTPTLQLPH 62
QY 62 AIEELFALAGKQLKDVVELLPVTFPRLCWESKVFYNDNDQAQLEAQIQGNPR----- 116
DB 53 LYKELFEENLNPADYVQLKELSPYTELK-----WDGTQDLTSDLSFKTQATLRL 115
QY 117 -DV-----AGYRAFL-DYSRAVFNEGYKLG-TVPFLSPKMLRAAPQL 157
DB 116 SDDLPAFDRWYSEHIRKYELGYPKLAGPARSPF--CYLRPDELMEKFLSPF----- 164
QY 158 AKLQAWRSVYKVGXIEDSHLRQAPSFHSLLVGGNPATTSIYTLIHALRENGVWYEPFR 217
165 -----PWNELTQHFRFPQDERLVYALSTPSKYLGMFTVASSVSLIPFLFSQGVVHPV 220
QY 218 GGTGALVNGMIKLPQDLGGEVVLNARVSHMETVGGKIQAQVLEDDGRRFETCAVASNADVV 277
DB 221 GGFALAQGLANAAQDLGVKIHLSHPVHQIWIQEQVGLLEADASRHPQDVTVINADFA 280
QY 278 HTYRDLLOH-PAKAKQAQKAKLOSXRMSNSFLVLYFGLNHHHDQLAHTVCFGRYRELIIH 337
DB 281 YAVRHLLPTSARGYTDNKLQGMQFSCSTFMLYGLNRRYEDLPHEQIYILSDNIRLERP 340
QY 338 IFNHDLGLAE-DFSILYHAPCVTDPSLAPEGCGSVYVLAIPVPHLGTANLDWAVEGFLRD 396
DB 341 WYDSDALDETDPPYVNCNPTIIDPSNAPAGHSTLYLVPTNTSYA-VDWIDIKQKSYTDF 399
QY 397 IFDYLEQHYMPGLARSOLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKH 456

b 400 ILKRLHLGYNIEQHIVTQSCYTAQSWLDDYRVHLGAVPNLSHNLTLQGPFRPIRSEN 459
Y 457 IDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLED 490
b 460 IAGLYWIGGAVHPGSGLLTILEASRSAGFIHQD 493

earch completed: February 29, 2004, 14:52:52
ob time : 15.6946 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:51:24 ; Search time 32.666 Seconds
(without alignments)
3180.293 Million cell updates/sec

tle: US-09-941-947A-32

irect score: 2606
quence: 1 MKPTTVIGAGFGGLAALRL.....IPGVIGSAKATAGLMLLEDLI 492

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

al number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	2606	100.0	492	10	US-09-941-947A-32
2	2606	100.0	492	14	US-10-218-118-8
3	1802	89.1	491	15	US-10-369-493-234
4	1595	61.2	494	9	US-09-547-267-5
5	1595	61.2	494	10	US-09-920-923-4
6	1464	56.2	498	15	US-10-369-493-10579
7	1461	56.1	506	15	US-10-369-493-17830
8	1183.5	45.4	498	15	US-10-369-493-20438
9	1051	40.3	548	15	US-10-369-493-540
10	1014	38.9	498	15	US-10-369-493-19248
11	962.5	36.9	518	15	US-10-369-493-7750
12	793.5	30.4	517	15	US-10-369-493-19519
13	703.5	27.0	511	9	US-09-934-903-16
14	703.5	27.0	511	9	US-09-934-868-76
15	703.5	27.0	511	10	US-09-941-947A-22

Sequence 4, Appli
Sequence 18, Appl
Sequence 78, Appl
Sequence 24, Appl
Sequence 8, Appl
Sequence 8951, Ap
Sequence 12, Appl
Sequence 8562, Ap
Sequence 8142, Ap
Sequence 18983, A
Sequence 3769, Ap
Sequence 21541, A
Sequence 18644, A
Sequence 14, Appl
Sequence 18, Appl
Sequence 19547, A
Sequence 7895, Ap
Sequence 3, Appl
Sequence 18646, A
Sequence 10196, A
Sequence 5696, Ap
Sequence 4194, Ap
Sequence 19518, A
Sequence 20440, A
Sequence 19986, A
Sequence 5, Appl
Sequence 6, Appl
Sequence 2831, Ap
Sequence 18878, A
Sequence 17259, A

511 14 US-10-358-917-4
497 9 US-09-934-903-18
497 9 US-09-934-868-78
497 10 US-09-941-947A-24
497 14 US-10-358-917-8
494 15 US-10-369-493-8951
502 14 US-10-358-917-12
513 14 US-10-156-761-8562
506 15 US-10-369-493-8142
499 15 US-10-369-493-18983
604 23.6
602 23.1
514 15 US-10-369-493-3769
512 15 US-10-369-493-21541
512 15 US-10-369-493-18644
497 14 US-10-358-917-14
530 14 US-10-128-713A-18
404 15 US-10-369-493-19547
495 15 US-10-369-493-7895
1268 13 US-10-438-784-3
536 15 US-10-369-493-18646
498 15 US-10-369-493-10196
439 9 US-09-815-242-5696
548 9 US-09-738-626-4194
485 15 US-10-369-493-19518
496 15 US-10-369-493-20440
537 15 US-10-369-493-19986
469 15 US-10-438-784-5
542 14 US-10-209-372-6
542 15 US-10-369-493-2831
539 15 US-10-369-493-18878
498 15 US-10-369-493-17259

ALIGNMENTS

RESULT 1

US-09-941-947A-32
; Sequence 32, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: C11903 US NA
; CURRENT APPLICATION NUMBER: US/09/941.947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-32

Query Match 100.0%; Score 2606; DB 10; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.2e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKPTTVIGAGFGGLAALRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVTIDP 60
Db 1 MKPTTVIGAGFGGLAALRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVTIDP 60
Qy 61 SAIEELPALAGKQLKDYVELLPVTFPRFLCWESGKVFNTNDQAQLAQIQFNPDRVAG 120

db 61 SAIEELFALAGKOLKDYVLLPVTPTFRLCWESGKVFYNDQAEQIQCFNPRDVAG 120
121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPEGGCGSYVVLAPVPHLGTANLDWAVEGPRLDRI FDYLEQHYMPGLRSOLVTHRMFT 420
361 SLAPEGGCGSYVVLAPVPHLGTANLDWAVEGPRLDRI FDYLEQHYMPGLRSOLVTHRMFT 420
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

ESULT 2
S-10-218-118-8
Sequence 8, Application US/10218118
Publication No. US20030148319A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: CUI876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: 60/312,646
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 492
TYPE: PRT
ORGANISM: Pantoea stewartii
US-10-218-118-8

Query Match 100.0%; Score 2506; DB 14; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.2e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
61 SAIEELFALAGKOLKDYVLLPVTPTFRLCWESGKVFYNDQAEQIQCFNPRDVAG 120
61 SAIEELFALAGKOLKDYVLLPVTPTFRLCWESGKVFYNDQAEQIQCFNPRDVAG 120
121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240

db 181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPEGGCGSYVVLAPVPHLGTANLDWAVEGPRLDRI FDYLEQHYMPGLRSOLVTHRMFT 420
361 SLAPEGGCGSYVVLAPVPHLGTANLDWAVEGPRLDRI FDYLEQHYMPGLRSOLVTHRMFT 420
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 3
US-10-369-493-234
Sequence 234, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 234
LENGTH: 491
TYPE: PRT
ORGANISM: Xenorhabdus nematophilus
US-10-369-493-234

Query Match 59.1%; Score 1802; DB 15; Length 491;
Best Local Similarity 67.4%; Pred. No. 4.5e-173;
Matches 331; Conservative 56; Mismatches 94; Indels 0; Gaps 0;

QY 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
DB 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
QY 61 SAIEELFALAGKOLKDYVLLPVTPTFRLCWESGKVFYNDQAEQIQCFNPRDVAG 120
DB 61 SAIEELFALAGKOLKDYVLLPVTPTFRLCWESGKVFYNDQAEQIQCFNPRDVAG 120
QY 121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
DB 121 YRAFLDYSAVFNVEGYLKIAGTVFPLSFKDMRAAPOLAKIQAWRSVYKVGAGYIDEHLR 180
QY 181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFLSHLLVGGNPFATSSYITLIHALEREWGVWFPFGGTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
DB 241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSHAPAAKQAKKLOSK 300
QY 301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
DB 301 RMNSLFLVYFGLNHHDDQLAHTVCFGPRYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360

3 301 KMSNSLFIYFGLNKLHICLAHTTICFGRNRYKELLDDIFEGNSIPGSPSLYLHSPSVTDP 360
/ 361 SLAPGCGSYYVLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
3 361 TLAPGCGSYYVLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
/ 421 PDFRDELNAMOGSAFSEVPILITQSARFPHNRDKHIDNLVVGAGTHPGAGIPGVIGSA 480
3 421 PDFRDELNAMOGSAFSEVPILITQSARFPHNRDKHIDNLVVGAGTHPGAGIPGVIGSA 480
/ 481 KATAGLMLEDL 491
3 481 KATAGLMLEDL 491

RESULT 4
1-09-547-267-5
Sequence 5, Application US/09547267
Patent No. US20020147371A1
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547.267
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
1-09-547-267-5
Query Match 61.2%; Score 1595; DB 9; Length 494;
Best Local Similarity 61.9%; Pred. No. 3.9e-152;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVICAGFGGLALAIRLOQAAGIPVLLLEQORDKPGRAYVYQEGFTFDAGPTVTITDP 60
1 MSSAIVICAGFGGLALAIRLOQAAGIATTTVEARDKPGRAYVWVNDQGHVFDAGPTVTITDP 60
61 SAIELEFALAGKQLKDYVELLPVTTPFYRLCWESCKVFYNDQALQLEAQIQOENPRDVAG 120
61 DSLRELWALSQPMERDVTLTPVSPFYRLTWADGRSEFYVNDDELIRQVASFNPADVVG 120

QY 121 YRAFIDYSRAVENEGYKLGITVPFLSPKDMLEAARAPQAKLQAMRSVYKVGAGYIBDEHLR 180
Db 121 YRFDYAEVYREGYKLGITVPFLKQGLMLNAPALMRLOAYRSVHSMWARFIQDPHLR 180
QY 181 QAFSPHLLVGCNPFATSSIVTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
Db 181 QAFSPHLLVGCNPFATSSIVTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVDKIQAVOLEDRGRFFETCAVASNADVVHTYRDLLSQHPAAAKQAKKIQSK 300
Db 241 NARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTTRGRRTKAILNRQ 300
QY 301 RMSNSLFVLYFGLNKHHDQLAHHTVCPGPRYRELTHETFNEDGLAEDPSLYLHAPCVTDP 360
Db 301 RMSNSLFVLYFGLNKHHDQLAHHTVCPGPRYRELTHETFNEDGLAEDPSLYLHAPCVTDP 360
QY 361 SLAPGCGSYYVLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
Db 361 SLAPGCGSYYVLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
QY 421 PDFRDELNAMOGSAFSEVPILITQSARFPHNRDKHIDNLVVGAGTHPGAGIPGVIGSA 480
Db 421 PDFRDELNAMOGSAFSEVPILITQSARFPHNRDKHIDNLVVGAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDL 491
Db 481 KATAGLMLEDL 491
RESULT 5
US-09-920-923-4
Sequence 4, Application US/09920923
Publication No. US20030022273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923-4
Query Match 61.2%; Score 1595; DB 10; Length 494;
Best Local Similarity 61.9%; Pred. No. 3.9e-152;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
QY 1 MKPTTVICAGFGGLALAIRLOQAAGIPVLLLEQORDKPGRAYVYQEGFTFDAGPTVTITDP 60
Db 1 MSSAIVICAGFGGLALAIRLOQAAGIATTTVEARDKPGRAYVWVNDQGHVFDAGPTVTITDP 60
QY 61 SAIELEFALAGKQLKDYVELLPVTTPFYRLCWESCKVFYNDQALQLEAQIQOENPRDVAG 120
Db 61 DSLRELWALSQPMERDVTLTPVSPFYRLTWADGRSEFYVNDDELIRQVASFNPADVVG 120
QY 121 YRAFIDYSRAVENEGYKLGITVPFLSPKDMLEAARAPQAKLQAMRSVYKVGAGYIBDEHLR 180
Db 121 YRFDYAEVYREGYKLGITVPFLKQGLMLNAPALMRLOAYRSVHSMWARFIQDPHLR 180
QY 181 QAFSPHLLVGCNPFATSSIVTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
Db 181 QAFSPHLLVGCNPFATSSIVTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVDKIQAVOLEDRGRFFETCAVASNADVVHTYRDLLSQHPAAAKQAKKIQSK 300
Db 241 NARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTTRGRRTKAILNRQ 300

301 RMSNSLVYFGLNHHDDQLAHTVCFGRPYRELTHEIFNHDGLADFSLYLHAPCVITDP 360
301 RWSNSLVFHLGSKRENLAHNSVIFGPYKGLVNEIFNGPRLPDGFSNLYHSPCVITDP 360
361 SLAPEGCGSVYVLAPVPHLGTANLWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
361 SLAPEGNSTHYVLAPVPHLGSADVDEAEPGYAEHIFELERRAIPDLRKLTVSRIFS 420
421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
421 PADFTSTLSAHHGSAFSEVPEILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPGVGSA 480
481 KATAGLMLEDL 491
481 KATAQWMLSDL 491

RESULT 6
S-10-369-493-10579
Sequence 10579, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10579
LENGTH: 498
TYPE: PRT
ORGANISM: Spingomonas aromaticivorans
S-10-369-493-10579

Query Match 56.2%; Score 1464; DB 15; Length 498;
Best Local Similarity 56.4%; Pred. No. 7e-139;
Matches 277; Conservative 78; Mismatches 136; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
7 MKRACVIGAGFGGLALAIRLQSGGVQTTLEARDKPGGRAYFMEKDGFTFDGPTVITDP 66
61 SAIEELFALAGKOLKDYVELLPVTPPYRLCWESGKVFNVDNDQALQLEAQIQCFNPRDVAG 120
67 ACLEELWALGHRMADVDVLPVMPYRLNWDGNTFDYSNDDSLRAEIAKLEPADVAG 126
121 YRAFLDYSRAVFNVEGYLKGTVPFPLSKDMLRAAPQALAKQAWRSYISKVAGYIEDEHLR 180
127 YDFLAYSAGVFEQYVKGNGVPLDFDASWIKAAALARYQAWRSYISWVSFVNEKLR 186
181 QAFSPHSLVGGNPPATSSITYLIHALEBWGWPPRGCTGALVNGMIKLFQDLGGEVVL 240
187 EAFSPHLLVGGNPMATSSIYALIHKLERDGGVWMAEGGTNRLIEGMVTHFRIGEVRI 246
241 NARVSHMETVGDKIQAQVLEDGRPRPTCAVSNADVHTYRLDLSCHPAAAKQAKLQSK 300
247 GQPVTEILTGDRVGVRTKSGVEVEDVAVSNADVHTYRLDLSCHPAAAKQAKLQSK 306
301 RMSNSLVYFGLNHHDDQLAHTVCFGRPYRELTHEIFNHDGLADFSLYLHAPCVITDP 360
307 RFSPLFVWHFEGTWPICPHENILFGPRYKGLLDDIYTHGVLPQDGSIVLHPTVITDP 366
361 SLAPEGCGSVYVLAPVPHLGTANLWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
367 SVAPEGNSTFYALVPVANMGKLPVNDEVGPILEKRLIDVEGRRRLIPDIHERIVTKFHYA 426

421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
427 PSDFSTLSAHHGSAFSEVPEILTQSAWFRPHNRDRAISNLYLVGAGTHPGAGIPGVGSA 486
481 KATAGLMLEDL 491
487 KATARLMLEDL 497

RESULT 7
US-10-369-493-17830
Sequence 17830, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 17830
LENGTH: 506
TYPE: PRT
ORGANISM: Spingomonas
US-10-369-493-17830

Query Match 56.1%; Score 1461; DB 15; Length 506;
Best Local Similarity 56.8%; Pred. No. 1.4e-138;
Matches 279; Conservative 75; Mismatches 137; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
15 MRAVWVIGAGFGGLALAIRLQSGAVDTTVEARDKPGGRAYVWERDGTFTDGGPTVITDP 74
61 SAIEELFALAGKOLKDYVELLPVTPPYRLCWESGKVFNVDNDQALQLEAQIQCFNPRDVAG 120
75 DAKELWRLSGHDISDVTLPKVPYRLSWPDGTTFDYTNDDAVLAGIAKLDSADTAD 134
121 YRAFLDYSRAVFNVEGYLKGTVPFPLSKDMLRAAPQALAKQAWRSYISKVAGYIEDEHLR 180
135 YRFLDYSASVVEGYVVLGAKAFLDQKSMKAAAPALMQHQAWSYISVSGFVNEKLR 194
181 QAFSPHSLVGGNPPATSSITYLIHALEBWGWPPRGCTGALVNGMIKLFQDLGGEVVL 240
195 EALSPTLLVGGNPMATSSIYALIHKLERDGGVWMAEGGTNRLIAGNITHFERLGGKVR 254
241 NARVSHMETVGDKIQAQVLEDGRPRPTCAVSNADVHTYRLDLSCHPAAAKQAKLQSK 300
255 GQPVTEILTGDRVGVRTKSGVEVEDVAVSNADVHTYRLDLSCHPAAAKQAKLQSK 314
301 RMSNSLVYFGLNHHDDQLAHTVCFGRPYRELTHEIFNHDGLADFSLYLHAPCVITDP 360
315 RFSPLFVWHFEGTWPICPHENILFGPRYKGLLDDIYTHGVLPQDGSIVLHPTVITDP 374
361 SLAPEGCGSVYVLAPVPHLGTANLWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
375 SVAPEGNSTFYALVPVANMGKLPVNDEVGPILEKRLIDVEGRRRLIPDIHERIVTKFHYA 434
421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
435 PTDPRDLAAHHGSGFSEVPEILTQSAWFRPHNRDRAISNLYLVGAGTHPGAGIPGVGSA 494
481 KATAGLMLEDL 491
495 KATAELMGL 505

RESULTS

USDA 8
-10-369-493-20438
Sequence 20438, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
US-10-369-493-540

9. FURTHER READING

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
Publication No. US20030233675A1
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493

```

; APPLICANT: Kinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTI
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 540
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Deinococcus radiodurans
US-10-369-493-540

Query Match 40.3%; Score 1051; DB 15; Length 548;
Best Local Similarity 43.5%; Pred. No. 4.8e-97;
Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9

QY 2 KPTTVIGAGFGGLAIRLQAGIPVLLRQRKPGGRAYVQ-EGQFTFDAGPTVITDP 60
DB 17 KTAIVIGAGFGGLGIRLQSLGFDITILRLDGPGRAYQKRTPGYVDFMGPTVITVP 76
QY 61 SAIEELPA-----LAGKQIKD-----YVELLPVTPPYRLCWES 93
DB 77 HFIEELPALERDRAGLDAPDYPPVLVSGREYKGVSGCPHTSYVTLVILPEYRIVFHD 136
QY 94 GKVFNYNDQALQLEBAQIQFNPDRVAGYRAPFLQYSRAVFNQGYLKLGTVPFLSFQDMRLA 153
DB 137 GTYFDYDGPSTRRQIAELAPGLAGYERFHADEAI FRGFLGLGYTHFGDVPTMLRV 196
QY 154 APQALAKQAWRSVYSKVANGYIEDHLRQAFSLVLGQNPRTSSIIYLIHALEREWCV 213
DB 197 VEDLLKLDVATLTSFTSKYFQSDKLQVSPFTLLVGGNPLSPVAYAMIHVFKTNGI 256
QY 214 MFRPGGTGALVNGMIKLFQDLGGEVVLINARVSMETVGD-----KIQAQVLEDGRRFETC 268
DB 257 HYAMGGTGALVRGLVQKFEELOGAIRYAGVDEVLDGNLPGKRTARGVRLSSGELRAD 316
QY 269 AVASNADVVHTYRDLLSQHPAAKQAKKLOSKEMSN--SLFVLVYLCNLHHHD-QLAHHTV 325
DB 317 LVASNGDWNTY--LKRVPISALVNSDLRVKAASEMSLLVYFGRGGDDLPLKHHNI 374
QY 326 CFGPRYRELHIEFMDHDAEDFSLYIHAPCVTDPDLAPEGCGSYVLPVPHLGATNLD 385
DB 375 LLGPRYELALLSEIFGTRKLGEDFSQYLHVPTLTDPALAPAGHAAAYTLVPVPHNG-SGID 433
QY 386 WAVEGRLRDRIPOYLRQH-YMGLRSQLVTHRMKFTPDFRDELANWQGSASVEPIILTO 444
DB 434 WVEGPGFLAEALADIERGLITGLERUITHTEFFITPDYFAGTLOSILGNAGPBPRLVQ 493
QY 445 SAWFPRNRRKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLUMLED 490
DB 494 SAFFRPNRRSDLNFLNYLVGAGAPQAGTSPVYMSAKMTARLIAD 539

```

RESUMEN 10

US-10-369-493-19248
Sequence 19248, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 19246
LENGTH: 498
TYPE: PRT
ORGANISM: Myxococcus xanthus
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(498)
OTHER INFORMATION: unsure at all Xaa locations
S-10-369-493-19248

Query Match 38.9%; Score 1014; DB 15; Length 498;
Best Local Similarity 44.0%; Pred. No. 2,3e-93;
Matches 210; Conservative 80; Mismatches 185; Indels 2; Gaps 2;
Y 6 VIGAGFGGLALAIRLQAAGIPVLLLEQKDPGRAYVVOEQGFTTDAAGTPTTDSATIEE 65
b 8 VIGSGFGGLAAATRLAARGWRVTVLERMDSPGRANAFOQDGTFTDAGTPTTCHLLEE 67
Y 66 LPAAGKQKLDVVELLPVTPFYRLCWESGKVFYNDQAOLEAQIQOQFNPRDVAGYRAFL 125
b 68 LNALGQQRADHVELPVPALYRMPEPDGSTDYHTDREAMQSVRLSPRDEAGFLALC 127
Y 126 DYSRAVFNEGYLKLTGVPFLSFQDMRLAAPQAKIQAQMSVYSKVAGYIEDEHLRQAFSP 185
b 128 ARVERMYEAGIGLMTSTPVDVLSAPPTPALVRDEAFMSGLVSKTIRDERLQOALS 187
Y 186 HSLLVGSGNPAT--SSYVTLIHALEREWGVPFGGTGALVNGMIKLFODLGGVVLNARV 244
b 188 HPLLVGSGFPPTASVTSIQFVERWAGFPVGGTGAALVRLVLLLESGLGEVRYGSEV 247
Y 245 SHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLLSQHPAAAKQAQKLSKRMNS 304
b 248 TEALGKRGATGVLGDTGTLAADVVSNAADAAVTVLPVGEVKKHTDREINRYSM 307
Y 305 SLVLYEGLNHHDDLAHTVCFGRPYRLIHEIFNHDGLAEDPSLYLHAPCVTDPSLAP 364
b 308 SAFELWYFGTRQVPEVAHHTLFGKDFGWFSGLEGQGPSADPLLYLHRPTATDALAP 367
Y 365 EGCGSYVVLAPVPHLGTANLDMWAVEGRFLRDRIFDYLEQHYMPGLRSQLVTHRMFTPDF 424
b 368 SGHDAFVVLAPVPHLGT-CAEWKQRAEAFRELEARSLSHTVLPGLGSELMTSRMATPEYF 426
Y 425 RDEINAWQGSAPSVBEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAK 481
b 427 RDELSRFKGAAPSFAPTLAQTTFLRAQARSEDVDCLYMWGAGTHPGAGLPAVLCSAK 483

RESULT 11
S-10-369-493-7750
Sequence 7750, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7750
LENGTH: 518
TYPE: PRT
ORGANISM: Rhodobacter sphaeroides

US-10-369-493-7750
Query Match 36.9%; Score 962.5; DB 15; Length 518;
Best Local Similarity 41.2%; Pred. No. 3,9e-88;
Matches 200; Conservative 78; Mismatches 201; Indels 7; Gaps 5;
QY 6 VIGAGFGGLALAIRLQAAGIPVLLLEQKDPGRAYVVOEQGFTTDAAGTPTTDSATIEE 65
Db 15 VIGSGFGGLAAATRLAARGWRVTVDKLDVPGGSGSSITQEGHREFDLGFTTIVTQSLRD 74
QY 66 LPAAGKQKLDVVELLPVTPFYRLCWESGKVFYNDQAOLEAQIQOQFNPRDVAGYRAFL 125
Db 75 LMTGCRDFDADVVELAPIDPFTFVWRPDGSHTVTQSTEAKEAEVARLSFGVAGYEXEL 134
QY 126 DYSRAVFNEGYLKLTGVPFLSFQDMRLAAPQAKIQAQMSVYSKVAGYIEDEHLRQAFSP 185
Db 135 KDSEKRYWFGYEDLGRSSMEKLDWLIKVLPTFGMRADRSVYQHAALRVKDRRLMALSF 194
QY 186 HSLLVGSGNPATSSYVTLIHALEREWGVPFGGTGALVNGMIKLFODLGGVVLNARVS 245
Db 195 HPLFVGDPFNTVSMYLVLSQLEKSPGVHYAIGGVAATAAAMAKVIEGOGGSRMTVEVD 254
QY 246 HMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLLSQHPAAAKQAQKLSKRMNS 305
Db 255 EILVEKGTATGVLASGEVLRAGLVVSNADAGHTYMRLLNENHPRRRTDARVKSRRWSMG 314
QY 306 LFWLYEGLNHH---DOLAHHTVCFGRPYRLIHEIFNHDGLAEDPSLYLHAPCVTDPSEL 362
Db 315 LFWYFGTGTGKMWPDVGHHTIVNAPRYKGLVEDIFLKGKLAQMSLYTHRPSITDPTV 374
QY 363 APGCGSYVVLAPVPHLGTAN-LDMWAVEGRFLRDRIFDYLEQHYMPGLRSQLVTHRMFTPT 421
Db 375 APGDDTFVALSPVPHLKAQOPVDMQVAEPYRESVLEVEQS-MFGIGERIGPSLVFTPT 433
QY 422 FDRDE-LNAWQGSAPSVBEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
Db 434 ETERDYLSPW-GAGFSIEPRILQSAWFRPENISEEVANFLVWAGTHPGAGVPGVIGSA 492
QY 481 KATAGL 486
Db 493 EVMAKL 498

RESULT 12
US-10-369-493-19519
Sequence 19519, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 19519
LENGTH: 517
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-10-369-493-19519
Query Match 30.4%; Score 793.5; DB 15; Length 517;
Best Local Similarity 36.1%; Pred. No. 4,8e-71;
Matches 175; Conservative 89; Mismatches 214; Indels 7; Gaps 4;
QY 6 VIGAGFGGLALAIRLQAAGIPVLLLEQKDPGRAYVVOEQGFTTDAAGTPTTDSATIEE 65

12 VVGUGVGLAAAAARLAHQGFVQVFEKQGPGRGNRLQVQDFTWDLGPTIVLMPVEFEE 71
66 LFLALAGKQLKDYVELLPVTPPYELLCWESGKRYFNVDNDQAQLEAQIQOFPNRPDVAGYRAEL 125
72 TFRVAGRRITEDYLLTLRCDPNRYVHFRRDGSWVTFSELCAWGRELERVEGCVYARVLAFL 131
126 DYSRVFNSEGYLKLGTVPFLSEFQDML--RAAPOLAKLOAMRSVYTSKVAGVIEDEHLRQAF 183
132 AQGRVQTRTSLDHLVGRNTAGVLDYLSPRVLARIFQVRAHRMTADVRSRFFQDERLRAAM 191
184 SFHSLLVGNPNPATSIYTLIIHALEREWGFWFPRGGTGALVNGMIKLFQDLGCGEVVLNAR 243
192 TFOQMYLVGVSYPASPAVYGLLPTELGVCIWPKGSLVAIPQALERLARERGVRFHYGAP 251
244 VSMETVGGKIQAVQLEDBGRRTCAVASNADVVHTYRDDLLSQHPAAKQAKKLQSKRHS 303
252 VERILTDGGRTRGVRLGSEGVVEADVLCNADLFPAYEKLLDPKATTLKRXEKL--RYT 308
304 NSLFLVLYFLGNHHDDQLAHTYVCFGRYRELIIHIFPNHDGLAEFSLYLHAPCVTPDPSLA 363
309 SSGYMLYLGMKRYPELLHNVVFGRDYKGSFDDIFERFVPEPDPFYVNAFTTQASLA 368
364 PEGCGSYVVLAVPHVLGHTANLDWAVEGPRLRDRIPDYLEQHVMPLGSLVQVTHRMTPTPD 423
369 PEGKDALYVLVPPH--QHPDLDWKVEGPKVRAKFFARWAEELGFPFLSEDSIEVERVEFTPD 427
424 FROELNAGWGSAPFSEPILTQSAWFRPNHRDKHDINLYLVGAGTHPCGAGIPGVIGSAK--A 482
428 WAGTFNLAGSAPGLSQNTQIGFPRPNQARVKNLFFVGASTQPTQGLPTVLEISARLV 487
483 TAGIAM 487
488 TERIAM 492

RESULT 13
 -09-934-903-16
 Sequence 16, Application US/09934903
 Patent No. US20020102690A1
 GENERAL INFORMATION:
 APPLICANT: Koffas, Mattheos
 APPLICANT: Odom, J. Martin
 APPLICANT: Schenzle, Andreas J.
 APPLICANT: No. US20020102690A1ton, Kelley C.
 APPLICANT: Tomb, Jean-Francois
 APPLICANT: Rouviere, Pierre
 APPLICANT: Picataggio, Stephen
 APPLICANT: Cheng, Giong
 TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
 FILE REFERENCE: CL1646 US NA
 CURRENT APPLICATION NUMBER: US/09/934, 903
 CURRENT FILING DATE: 2001-08-22
 PRIOR APPLICATION NUMBER: 60/229,907
 PRIOR FILING DATE: September 1, 2001
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 16
 LENGTH: 511
 TYPE: RPT
 ORGANISM: Methylobionas 16a
 FEATURE:
 OTHER INFORMATION: Amino acid sequences encoded by ORF8
 09-934-903-16

Query Match 27.0%; Score 703.5; DB 9; Length 511;
Best Local Similarity 31.7%; Pred.No. 66-62;
Matches 156; Conservative 105; Mismatches 210; Indels 25; Gaps 10

2 KPTVTGAGFGGIALAIRLQAAGIIVLLLEQRDPGGRVYVQEGSTPDAGTVIDFS 61
5 KHIIVGAGPGGICAGMLISQRFKVSIFDKHATIGGRNRINNNNGTFTGTFTLLMG 64
62 AISEPALAGKOLKDVELLPTVTPYFLCWESGKVFYNDQAGLEAOICQFNPRVAGY 121

Db	65	VLDENFELCRRSEYDEFLPLSPMYELLDDRDIFVY-SDRENMAEQLQRFVDEGTG	123
Qy	122	RAFLDYSAVAFNEG-----LKLGTVPFLSPFKMLRAAPQLAKLQAWRSVSVKAG	172
Db	124	EQPMQERKRFNALYPCITRDYSLK-----SPLSL-DLIKALPWLA---PPKSVFNLLQ	175
Qy	173	YIEDHLRAQPSFHSLLVGGNPFATSIYTLIRALERGVWVPPRGGTGAALVNGMIKLQ	232
Db	176	YFNOEKMRILAFQSKYLGNSPWPCALFTMLPYLEHEYGIYHVKGGLNRIAAAMAQVIA	235
Qy	233	DIGGEVVLNARVSHMETVGDKIQAVQLEDOGRRETCAVASNADVVTYRDDLSQHPAAK	292
Db	236	ENGGETHLSNRIESLIENGGAQGVKLQHGAEIRGDEVITINADFAHAMTLVK--PGVLK	293
Qy	293	Q--AKKLQSKMSNLSFLVYFGLNHHDLQAHETVCGFPYRELIIHEIFNHDGLAEDFSL	350
Db	294	KYTPENMLKORYSYSTFMLYGLDKIYD-LPHETIVFADYTNINRNIFDNKTLTDDPSF	352
Qy	351	YLKAPCVTDPSLAPEGCGSYVLAAPVPHLGTANLDNAVEGRDRIFFYLEQHV--MPGL	409
Db	353	YVONASADDSILAAGKASALYVIVPMEN-NDSGLDQWQHCVNEQVLTILGARLGSDI	411
Qy	410	RSQLYVTHRMFTFFDPRSELNAQGSASFVSEPIILITQSAWFRPHNRKXHDNLNLYLVAGTHT	469
Db	412	RAHICEKXIITPQWETDEHVYKCATFSLSHKFSQMLYWRPHNRFEELANCYLVGGGTHP	471
Qy	470	GAGIPGVIGSAKATAGLM	487
Db	472	GSGLPTIYESARISAKLI	489

```

RESULT 14
US-09-934-868-76
  Sequence 76, Application US/09934868
  Patent No. US20020137190A1
  GENERAL INFORMATION:
  APPLICANT: Koffas, Mattheos
  APPLICANT: Odum, James M
  APPLICANT: Schenzle, Andreas J
  TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
  FILE REFERENCE: CL1596 US NA
  CURRENT APPLICATION NUMBER: US/09/934,868
  CURRENT FILING DATE: 2001-08-22
  PRIOR APPLICATION NUMBER: 60/229,858
  PRIOR FILING DATE: 2000-09-01
  NUMBER OF SEQ ID NOS: 81
  SOFTWARE: Microsoft Office 97
  SEQ ID NO 76
  LENGTH: 511
  TYPE: PRT
  ORGANISM: Methylomonas 16a
  FEATURE:
  OTHER INFORMATION: Amino acid sequences encoded by CRTN1
US-09-934-868-76

```

Query Match	27.0%;	Score	703.5;	DB	9;	Length	511;
Best Local Similarity	31.7%;	Pred.	No. 6e-62;				
Matches	158;	Conservative	105;	Mismatches	210;	Indels	25;
Gaps	10;						
Qy	2	KETTIVIGAGPGLALATRLQAAGTFLVLLSORDKPGRAVVTQEQGTFDAGCTVTITDPS	61				
Db	5	KHIIIVGAPGLCAGMLLSORGFKVISIPKHAIEIGRRRPFNNMGFTDTGTFELLMKG	64				
Qy	62	AIEELFALGAKQLKDYVELLPVTFYRLCWESKVNRYNDQAQLEAQIQQNPRDVAGY	121				
Db	65	VLDENFELCERESDYLEFPLSPMYRLLYDDRDIFVY-SDRENKAEALQRFVDEGTGQY	123				
Qy	122	RAFLDYSRAVNEG------LKLGTVPTLSPKDMLRAPOLAKLQANRSVTSKVG	172				
Db	124	EQFMQERRRNALYPCITRDYSLLK-----SFLSL-DLIKALPWLA---PSPKSVNNLQQ	175				
Qy	173	YIEDEHLRQAFSHSLVGGNPNPATISSYITLIHALREWGVNPPRGGTGALVNGMIKLQF	232				

```

b 176 YNQRRLAFQSKYLGMSWSPALFTMLPYLHEHYIYHVKGGLNRFAAQAQVIA 235
y 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRFPETCAVASNADVVHTYRDLLSQHFAAAK 292
b 236 ENGGEHLNSELIESLIIENGAAGKVKLQGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
y 293 Q--AKKLOSKRMSNSLFVLYFGLNHHHDLQAHHTVCFGPYRELHIEIFNHDGLAEDFSL 350
b 294 KYPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTTNIRNIFDNKTITDDFSP 352
y 351 YLHAPCVTPDPSLAPEGCGSYVYVLAAPVPHLGTANLMDWAVEGPRLEDRIFDYLEQHY-MPGL 409
b 353 YQNASASDDSLAPAGKSALYVLPMPEN-NDSGLDQAHQCNVREQVLDTLGARGLSDI 411
y 410 RSLVTHRMFTFPDFDELNANQGSAPSVPEPILTQSAMFRPHNRDKHIDNLYLVGAGTHP 469
b 412 RAHIECKEITPTQWETDEHVYKGFATFSLSHKFSQMLYWRPHNRFEEELANCYLVGCGTHP 471
y 470 GAGIPGVIGSAKATAGLM 487
b 472 GSGLPTIYESARISAKLI 489

```

RESULT 15

S-09-941-947A-22
Sequence 22, Application US/09941947A
Publication No. US20030003528A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: DiCosimo, Deana J.
APPLICANT: Koffas, Mattheos
APPLICANT: Miller, Edward S. Jr.
APPLICANT: Odum, J. Martin
APPLICANT: Picataggio, Steve
APPLICANT: Rouviere, Pierre E.
TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
FILE REFERENCE: CL1903 US NA
CURRENT APPLICATION NUMBER: US/09/941,947A
CURRENT FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,858
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Microsoft Office 97
SEQ ID NO 22
LENGTH: 511
TYPE: PRT
ORGANISM: Methylobionas 16a
IS-09-941-947A-22

Query Match 27.0%; Score 703.5; DB 10; Length 511;
Best Local Similarity 31.7%; Pred. No. 6e-62;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

```

y 2 KPTTVIGAGGGALAIRLOAGIPVLLLEQORDKPGRAVYVYQEGFTYDAGTWTIDPS 61
b 5 KHIIVGAGPGGLCAGKLLSQRGFKVSIFDKHAEIGGRNRPINMNGFTFTGTFTILMKG 64
y 62 AIELFALAGKQLKDYVELLPVTPFYRLCWESKGVFNVDNDQALEAQIQOFPNPRDVAGY 121
b 65 VLDEMFELCERSEDIYLFPLSPMYELLVYDDRDIFVY-SDRENWGAELQRVFEDETDGY 123
y 122 RAFLDYGRAVFNQY-----LKIGTVPLSFYKMDLRAAPOLAKLQAWRSVYKVG 172
b 124 EQFMEQERKRFNALYPCITRDYSLK----SFLSL-DLIALPWLAA--FPKSVFNLLQ 175
y 173 YIEDEHLRQAFSFLVGVGNPPATSIYTLIALEREMGWFPFGTGTGALVNGMIKLPQ 232
b 176 YFNQEKMLAFQCFQSKYLGMSWSPALFTMLPYLHEHYIYHVKGGLNRFAAQAQVIA 235

```

```

Qy 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRFPETCAVASNADVVHTYRDLLSQHFAAAK 292
Db 236 ENGGEHLNSELIESLIIENGAAGKVKLQGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
Qy 293 Q--AKKLOSKRMSNSLFVLYFGLNHHHDLQAHHTVCFGPYRELHIEIFNHDGLAEDFSL 350
Db 294 KYPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTTNIRNIFDNKTITDDFSP 352
Qy 351 YLHAPCVTPDPSLAPEGCGSYVYVLAAPVPHLGTANLMDWAVEGPRLEDRIFDYLEQHY-MPGL 409
Db 353 YQNASASDDSLAPAGKSALYVLPMPEN-NDSGLDQAHQCNVREQVLDTLGARGLSDI 411
Qy 410 RSLVTHRMFTFPDFDELNANQGSAPSVPEPILTQSAMFRPHNRDKHIDNLYLVGAGTHP 469
Db 412 RAHIECKEITPTQWETDEHVYKGFATFSLSHKFSQMLYWRPHNRFEEELANCYLVGCGTHP 471
Qy 470 GAGIPGVIGSAKATAGLM 487
Db 472 GSGLPTIYESARISAKLI 489

```

Search completed: February 29, 2004, 15:28:02
Job time : 34.666 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 14.951 Seconds
(without alignments)
1698.885 Million cell updates/sec

US-09-941-947A-32

fect score: 2806

quence: 1 MKPTTVIGAGFGGLAIAIRL.....IPGVIGSAKATAGLMLELII 492

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgm2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgm2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgm2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgm2_6/ptodata/2/iaa/6CTUS_COMB.pep.*
6: /cgm2_6/ptodata/2/iaa/6CTUS_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2438	93.6	492	1 US-07-783-705A-4	Sequence 4, Appli
2	1849.5	71.0	489	1 US-08-095-726-8	Sequence 8, Appli
3	1849.5	71.0	489	1 US-08-096-043-8	Sequence 8, Appli
4	1849.5	71.0	489	1 US-08-096-623A-8	Sequence 10, Appl
5	1845.5	70.8	489	1 US-08-095-726-10	Sequence 10, Appl
6	1845.5	70.8	489	1 US-08-096-043-10	Sequence 10, Appl
7	1845.5	70.8	489	1 US-08-096-623A-10	Sequence 10, Appl
8	1595	61.2	494	3 US-08-660-645A-5	Sequence 5, Appli
9	1595	61.2	494	3 US-09-298-718-5	Sequence 5, Appli
10	1595	61.2	494	3 US-09-548-969-5	Sequence 5, Appli
11	1595	61.2	494	3 US-08-980-832-4	Sequence 4, Appli
12	1595	61.2	494	4 US-09-547-267-5	Sequence 4, Appli
13	1595	61.2	494	4 US-09-920-923B-4	Sequence 4, Appli
14	703.5	27.0	511	4 US-09-934-903-16	Sequence 16, Appl
15	687	26.4	497	4 US-09-934-903-18	Sequence 18, Appl
16	582	21.6	582	4 US-09-091-725-17	Sequence 17, Appl
17	259.5	10.0	290	4 US-08-936-165A-462	Sequence 462, App
18	201	7.7	610	4 US-09-443-184-55	Sequence 55, Appl
19	140	5.4	527	1 US-08-132-168A-32	Sequence 32, Appl
20	134	5.1	544	1 US-08-472-028A-8	Sequence 8, Appli
21	134	5.1	544	2 US-08-808-931-8	Sequence 8, Appli
22	134	5.1	544	3 US-08-808-323-8	Sequence 8, Appli
23	134	5.1	544	3 US-09-050-603A-8	Sequence 8, Appli
24	134	5.1	544	3 US-09-102-420B-8	Sequence 8, Appli
25	134	5.1	544	3 US-09-071-296-8	Sequence 8, Appli
26	134	5.1	544	3 US-09-196-268-8	Sequence 8, Appli
27	134	5.1	544	3 US-09-015-683-8	Sequence 8, Appli

28	134	5.1	544	4 US-09-191-998-8	Sequence 8, Appli
29	134	5.1	544	4 US-09-437-928-8	Sequence 8, Appli
30	131	5.0	458	4 US-09-328-352-7549	Sequence 7549, Ap
31	130	5.0	414	3 US-09-067-626-4	Sequence 4, Appli
32	121.5	4.7	571	4 US-09-690-942-15	Sequence 15, Appl
33	118	4.5	409	4 US-09-252-991A-24944	Sequence 24944, A
34	118	4.5	474	4 US-09-543-681A-6873	Sequence 6873, Ap
35	118	4.5	614	4 US-09-567-003C-22	Sequence 22, Appl
36	116.5	4.5	524	3 US-09-210-388-1	Sequence 1, Appli
37	116.5	4.5	524	3 US-09-912-176-1	Sequence 1, Appli
38	115.5	4.4	578	4 US-09-690-942-10	Sequence 10, Appl
39	115	4.4	508	1 US-08-472-028A-4	Sequence 4, Appli
40	115	4.4	508	2 US-08-808-931-4	Sequence 4, Appli
41	115	4.4	508	3 US-08-808-323-4	Sequence 4, Appli
42	115	4.4	508	3 US-09-050-603A-4	Sequence 4, Appli
43	115	4.4	508	3 US-09-102-420B-4	Sequence 4, Appli
44	115	4.4	508	3 US-09-071-296-4	Sequence 4, Appli
45	115	4.4	508	3 US-09-196-268-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-07-783-705A-4
; Sequence 4, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeyuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 13-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-4

Query Match 93.6%; Score 2438; DB 1; Length 492;

Best Local Similarity 93.5%; Pred. No. 1e-234;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDGGRAYVYQSGFTFDAGPTVITDP 60
1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDGGRAYVYQSGFTFDAGPTVITDP 60

61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFYNDQLEAQIQOQFNPRDVAG 120
61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFYNDQLEAQIQOQFNPRDVAG 120

121 YRAFLDYSRVAFNEGKLVKGTVPFLSFKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
121 YRAFLDYSRVAFNEGKLVKGTVPFLSFKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180

181 QAFSFLSLVGGNPPATSSITLILHALEBWGVWPPRGGTGALVNGMIKLFODLGGEVVL 240
181 QAFSFLSLVGGNPPATSSITLILHALEBWGVWPPRGGTGALVNGMIKLFODLGGEVVL 240

241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLLSOHAAAQAKKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLLSOHAAAQAKKLOSK 300

301 RMNSLFLVYFGLNHHHDLAHTVCGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360
301 RMNSLFLVYFGLNHHHDLAHTVCGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLEOHYMPGLRSQVLTHERMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLEOHYMPGLRSQVLTHERMFT 420

421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVAGTHPGAGIPGVIGSA 480
421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVAGTHPGAGIPGVIGSA 480

481 KATAGLMEDLI 492
481 KATAGLMEDLI 492

RESULT 2
US-08-095-726-8
Sequence 8, Application US/08095726
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELEPHONE: 31285667180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-8

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
Best Local Similarity 72.3%; Pred. No. 6.1e-176;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDGGRAYVYQSGFTFDAGPTVITDP 60
DB 1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDGGRAYVYQSGFTFDAGPTVITDP 60

QY 61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFYNDQLEAQIQOQFNPRDVAG 120
DB 61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFYNDQLEAQIQOQFNPRDVAG 120

QY 121 YRAFLDYSRVAFNEGKLVKGTVPFLSFKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
DB 121 YRAFLDYSRVAFNEGKLVKGTVPFLSFKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180

QY 181 QAFSFLSLVGGNPPATSSITLILHALEBWGVWPPRGGTGALVNGMIKLFODLGGEVVL 240
DB 181 QAFSFLSLVGGNPPATSSITLILHALEBWGVWPPRGGTGALVNGMIKLFODLGGEVVL 240

QY 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLLSOHAAAQAKKLOSK 300
DB 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLLSOHAAAQAKKLOSK 300

QY 301 RMNSLFLVYFGLNHHHDLAHTVCGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360
DB 301 RMNSLFLVYFGLNHHHDLAHTVCGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360

QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLEOHYMPGLRSQVLTHERMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLEOHYMPGLRSQVLTHERMFT 420

QY 421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVAGTHPGAGIPGVIGSA 480

QY 481 KATAGLMEDLI 492
DB 481 KATAGLMEDLI 492

RESULT 3
US-08-096-043-8
Sequence 8, Application US/08096043
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL

COUNTRY: USA
 ZIP: 60680-0703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096.043
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,568
 FILING DATE: 30-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5530189val B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO. 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 :08-096-043-8

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
 Best Local Similarity 72.3%; Pred. No. 6.1e-176;
 Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIAGFGGLALALRLQAAGIPVLLLEQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
 1 MKPTTVIAGFGGLALALRLQAAGIPVLLLEQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
 61 SAIEELPALAGKQKDYVELLPVTPFYFLCHESKVFYNDQALQAEQIQFPRVAG 120
 61 TALEALFTLAGRMEDYVRLLPVKPFYFLCHESGKTLDYANDSFELEAQITQFPRVAG 120
 121 YRAFLDYGRAVNEGYLKGTVPFPLSKDMLRAAPOLAKLQAWSVYSKVAGYIEDEHLR 180
 121 YRFLAYSCAVFQEGYLRIGSVFPLSKDMLRAAGFOLLKQAWSVYSQVSSEFIEDEHLR 180
 181 QAFSPHSLVGNPNFATSIYTLIHALEREWGVWPPRGGTGALVNGMIKLPQDLGGEVVL 240

301 RMSNLSFLVLYFGLNEHDLQLAHHTVCFGRYRELTHIEFNHDLGAEPSLYLHAPCVTDP 360
 301 SMSNLSFLVLYFGLNPHSLQAHHTICFGRYRELIDEFTGSALADDFSLYLSHPCVTDP 360
 361 SLAPEGCSYVYLAIVPHGLTANLDAVEGPRLEDRIDPYLEQHYMPGLRSQVTHRMFT 420
 361 SLAPPPCASFYVLAIVPHGLNAPLDAWQEGPKLRIDRIPDYLEERYMPGLRSQVTRIFT 420
 421 PDFDFELNMGSAFSEPIILTSQAWFRPNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 421 ROTSRHMAIILGSLFIPEPPLSTQGLF--AANATRH--SNLYLVAGTHPGAGIPGVIGLA 477
 481 KATAGLMLEDL 491
 478 ESTASLMIEDL 488

35ULT 4
 3-08-096-623A-8
 Sequence 8, Application US/08096623A
 Patent No. 5684238
 GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
 APPLICANT: Brinkhaus, Friedhelm L.
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H.
 APPLICANT: Yarger, James G.
 APPLICANT: Yen, Hwei-Che B.
 TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
 TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 104
 CORRESPONDENCE ADDRESS:
 ADDRESSES: Welsh & Katz, Ltd.
 STREET: 120 S. Riverside Plaza, 22nd Floor
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096.623A
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,061
 FILING DATE: 09-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/662,921
 FILING DATE: 28-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/562,674
 FILING DATE: 03-AUG-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/525,551
 FILING DATE: 18-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,613
 FILING DATE: 02-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gamson, Edward P.
 REGISTRATION NUMBER: 29,381
 REFERENCE/DOCKET NUMBER: AMO-006.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 655-1500
 TELEFAX: (312) 655-1501
 INFORMATION FOR SEQ ID NO. 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-096-623A-8

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
 Best Local Similarity 72.3%; Pred. No. 6.1e-176;
 Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIAGFGGLALALRLQAAGIPVLLLEQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
 1 MKPTTVIAGFGGLALALRLQAAGIPVLLLEQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
 61 SAIEELPALAGKQKDYVELLPVTPFYFLCHESKVFYNDQALQAEQIQFPRVAG 120
 61 TALEALFTLAGRMEDYVRLLPVKPFYFLCHESGKTLDYANDSFELEAQITQFPRVAG 120
 121 YRAFLDYGRAVNEGYLKGTVPFPLSKDMLRAAPOLAKLQAWSVYSKVAGYIEDEHLR 180
 121 YRFLAYSCAVFQEGYLRIGSVFPLSKDMLRAAGFOLLKQAWSVYSQVSSEFIEDEHLR 180
 181 QAFSPHSLVGNPNFATSIYTLIHALEREWGVWPPRGGTGALVNGMIKLPQDLGGEVVL 240

181 QAFSHLLVGGNPF... 240
 241 NARVSHMETVGD... 300
 241 NARVEELVADNRV... 300
 301 RMSNSLVLYFGLNHHDDLAHTV... 360
 301 SMSNSLVLYFGLNQPHS... 360
 361 SLAPEGCGSVYVLA... 420
 361 SLAPPCASFVLA... 420
 421 PFDRDELNAWQGS... 480
 421 RQTSRHAWIALGSLF... 477
 481 KATAGLMEIDL 491
 478 ESTASLMEIDL 488

RESULT 5
 US-08-095-726-10
 Sequence 10, Application US/08095726
 Patent No. 5530188
 GENERAL INFORMATION:
 APPLICANT: Ausich, Rodney L
 APPLICANT: Brinkhaus, Friedhelm L
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H
 APPLICANT: Yarger, James G
 APPLICANT: Yen Hwei-Che B
 TITLE OF INVENTION: Beta-Carotene Biosynthesis in
 TITLE OF INVENTION: Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 79
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Amoco Corp., Patents and Licensing Dept
 STREET: 200 E Randolph St
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60680-0703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/095,726
 FILING DATE: 21-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,566
 FILING DATE: 30-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5530188val B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-095-726-10

Query Match 70.8%; Score 1845.5; DB 1; Length 489;
 Best Local Similarity 72.1%; Pred. No. 1:5e-175;
 Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTVTGAGCGGLALAIRQAAGIPVLLLEQRDKPGGRAYVYQEGTFDAGPTVITDP 60
 DB 1 MEKTVVIGAGFGGLALAIRQAAGIPVLLLEQRDKPGGRAYVYQEGTFDAGPTVITDP 60
 QY 61 SATBELPALAGKOLKDYVELLPVTPYRLCWESGKVFNTDNDQAQLEAQIQPNPRDVAG 120
 DB 61 TALEALFTLAGRRMEDVYRLLPVKPFYRLCWESGKTLDYANDSFELEAQITQFNPRDVAG 120
 QY 121 YRAPLDYSRAVENEGYKLGTVPLSPKDMRAAPOLAKLOAMRSVYKVAGYIEDEHLR 180
 DB 121 YRRPLAYSQAVFOEGYLRUGSVPLSPKDMRAAPOLAKLOAMRSVYKVAGYIEDEHLR 180
 QY 181 QAFSHLLVGGNPFATSIYTLIHALEREWGVPFRGGTGAIVNGMKLFDLGGEVVL 240
 DB 181 QAFSHLLVGGNPFATSIYTLIHALEREWGVPFRGGTGAIVNGMKLFDLGGEVVL 240
 QY 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLISQHPAAAKQAKLOSK 300
 DB 241 NARVEELVADNRVSVQRLADGRIFDQDAVASNADVVHTYRDLISQHPAAAKQAKLOSK 300
 QY 301 RMSNSLVLYFGLNHHDDLAHTVCGPPRYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 360
 DB 301 SMSNSLVLYFGLNQPHSQAHTICGPPRYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 360
 QY 361 SLAPEGCGSVYVLAAPVPHLGTANLDNAVEGPRLRDRIFDYLEQHYMGLRSOLVTHRMFT 420
 DB 361 SLAPPCASFVLAAPVPHLGTANLDNAVEGPRLRDRIFDYLEQHYMGLRSOLVTHRMFT 420
 QY 421 PFDRDELNAWQGSFVEPILITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 DB 421 RQTSRHAWIALGSLFIEPSPSLTQGLF--AANATRH-SNLYLVAAGTHPGAGIPGVIGSA 477
 QY 481 KATAGLMEIDL 491
 DB 478 ESTASLMEIDL 488

RESULT 6
 US-08-096-043-10
 Sequence 10, Application US/08096043
 Patent No. 5530189
 GENERAL INFORMATION:
 APPLICANT: Ausich, Rodney L
 APPLICANT: Brinkhaus, Friedhelm L
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H
 APPLICANT: Yarger, James G
 APPLICANT: Yen Hwei-Che B
 TITLE OF INVENTION: Lycopen Biosynthesis in
 TITLE OF INVENTION: Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 70
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Amoco Corp., Patents and Licensing Dept
 STREET: 200 E Randolph St
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60680-0703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,043
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,566
 FILING DATE: 30-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5530189val B

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO: 10:
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,623A
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,061
 FILING DATE: 09-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/662,921
 FILING DATE: 28-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/562,674
 FILING DATE: 03-AUG-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/525,551
 FILING DATE: 18-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,613
 FILING DATE: 02-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gamson, Edward P.
 REGISTRATION NUMBER: 29,381
 REFERENCE/DOCKET NUMBER: AMO-006.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 655-1500
 TELEFAX: (312) 655-1501
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 1-08-096-043-10

Query Match 70.8%; Score 1845.5; DB 1; Length 489;
 Best Local Similarity 72.1%; Pred. No. 1.5e-175;
 Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDP 60
 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDP 60
 61 SAIBELFALAGKQKDYVELLPVTFYELCWESGKVFNVDNDQALQAEQAOQFNPRDVAG 120
 61 TALEALFTLAGRMEDYVRLPVKPYELCWESGKTYLDYANDSFELEAQITQFNPRDVAG 120
 121 YRAFLDYRAVFNEGYLKIGTVPFSLFKDMLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
 121 YRRFLAYSQAVFQEGYLRGLSVFSLFRDMLRAGPQLLKQAWQSVYQSVSRFIEDEHLR 180
 181 QAFSHSLVGNPNPATSSIYTLIHALREWGVPFGGTGALVNGMKLPQDLGGEVVL 240
 181 QAFSHSLVGNPNPTTSIYTLIHALREWGVPFGGTGALVNGMKLPQDLGGEVVL 240
 241 NARVSHMETVGDKIQAQVLEDDRRFETCAVASNADVVHTYRDLLSQHPAAKQAKLQSK 300
 241 NARVVELVADNRVSVQRLADGRIFEDTAVASNADVNTYKKLLGTIPVQKRAARLERK 300
 301 RMSNSLVLYFGLNHHQDLAHTVCFGRPRELHETFNHDLGABDSLYLHAPCVTDP 360
 301 SMSNSLVLYFGLNQPHSQLAHHTICFGRPRELHETFTGSALADDFSLYHSPCVTDP 360
 361 SLAPECGSVYVYLAVPVPHLGTANLDWAVEGPELRDRIDYLEQHYMPGLRSQVTHRMET 420
 361 SLAPPCASFYVYLAVPVPHLGNAPLDWAQEGPKLRIDYLEERYPGLRSQVTHRIFT 420
 421 PFDFRDELNAWQGSAPFVEPILTSQSAWFRPNRDKHDNLVYVGAGTHPGAGIPGVIGSA 480
 421 RQTSRHWATILGSLFIBPSSLTQGLF--AANATRH-SNLVYVAAGTHPGAGIPGVVGLA 477
 481 KATAGLMEDL 491
 478 ESTASLMIEDL 488

RESULT 7
 S-08-096-623A-10
 Sequence 10, Application US/08096623A
 Patent No. 5684238
 GENERAL INFORMATION:
 APPLICANT: Ausich, Rodney L.
 APPLICANT: Brinkhaus, Friedrich L.
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H.
 APPLICANT: Yarger, James G.
 APPLICANT: Yen, Huel-Che B.
 TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
 TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 104
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Welsh & Katz, Ltd.
 STREET: 120 S. Riverside Plaza, 22nd Floor
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60606

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,623A
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,061
 FILING DATE: 09-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/662,921
 FILING DATE: 28-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/562,674
 FILING DATE: 03-AUG-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/525,551
 FILING DATE: 18-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,613
 FILING DATE: 02-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gamson, Edward P.
 REGISTRATION NUMBER: 29,381
 REFERENCE/DOCKET NUMBER: AMO-006.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 655-1500
 TELEFAX: (312) 655-1501
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 10-08-096-623A-10

Query Match 70.8%; Score 1845.5; DB 1; Length 489;
 Best Local Similarity 72.1%; Pred. No. 1.5e-175;
 Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDP 60
 DB 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDP 60
 QY 61 SAIBELFALAGKQKDYVELLPVTFYELCWESGKVFNVDNDQALQAEQAOQFNPRDVAG 120
 DB 61 TALEALFTLAGRMEDYVRLPVKPYELCWESGKTYLDYANDSFELEAQITQFNPRDVAG 120
 QY 121 YRAFLDYRAVFNEGYLKIGTVPFSLFKDMLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
 DB 121 YRRFLAYSQAVFQEGYLRGLSVFSLFRDMLRAGPQLLKQAWQSVYQSVSRFIEDEHLR 180
 QY 181 QAFSHSLVGNPNPATSSIYTLIHALREWGVPFGGTGALVNGMKLPQDLGGEVVL 240
 DB 181 QAFSHSLVGNPNPTTSIYTLIHALREWGVPFGGTGALVNGMKLPQDLGGEVVL 240
 QY 241 NARVSHMETVGDKIQAQVLEDDRRFETCAVASNADVVHTYRDLLSQHPAAKQAKLQSK 300
 DB 241 NARVVELVADNRVSVQRLADGRIFEDTAVASNADVNTYKKLLGTIPVQKRAARLERK 300
 QY 301 RMSNSLVLYFGLNHHQDLAHTVCFGRPRELHETFNHDLGABDSLYLHAPCVTDP 360
 DB 301 SMSNSLVLYFGLNQPHSQLAHHTICFGRPRELHETFTGSALADDFSLYHSPCVTDP 360
 QY 361 SLAPECGSVYVYLAVPVPHLGTANLDWAVEGPELRDRIDYLEQHYMPGLRSQVTHRMET 420
 DB 361 SLAPPCASFYVYLAVPVPHLGNAPLDWAQEGPKLRIDYLEERYPGLRSQVTHRIFT 420
 QY 421 PFDFRDELNAWQGSAPFVEPILTSQSAWFRPNRDKHDNLVYVGAGTHPGAGIPGVIGSA 480

421 ROTSRHAWIAILGSLFTEPESLTQGLF--AANATRH-SNLYVAAAGTHPGAGIPGVWGLA 477
481 KATAGLMEDL 491
478 ESTASLMIEDL 488

RESULT 8
S-08-660-645A-5
Sequence 5, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-660-645A-5

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEGFTFDAGPTVITDP 60
1 MSSAIVIGAGFGGLALAIRLQSAAGIATTVIARDKPGRAYVNDQGHVFDAGPTVITDP 60
61 SAIEELPALAGKQKDVVELLPVTPVELCWESGKVPYNDQAOLEAQIQPNRVDVAG 120
61 DSURELWALSGQMERDVTLLPSPFRLTWADGRSFYVNDDELIRQVASFNPADVDG 120
121 YRAFLDYSAVFNVEGYLKLGTVPFLSPKQMLRAAPQLAKLQAWRSYYSKVAGYIEDEHLR 180
121 YRRFDVAEYVREGYLYLKLGTFTFLKLGQMLNAPALMRQLQAVSVHSMVARFIQDPHLR 180
181 QASFSHLLVGGNPPATSSITYTLHLERWGVWPPGGTGGALVNGMKLFDQILGGEVVL 240
181 QAFSFTLLVGGNPPFSTSSITYTLHALERGGVWFARGGTGNQLVAGWALFRLGSLTL 240

QY 241 NARVSHWETVGDKIQAVQLEDGERFRFTCAVASNADVVHTYRDLLOHPPAAKQKQKQSK 300
DB 241 NARVTRIDTEGDRATGVTLDCRQLRADTVASNGDVMSYRDLGHTREGRTKAAILNEQ 300
QY 301 RMSNSFLVLYGLNHEHDLAHHHTVCGPRYRELIIHEIFNHDGLAEDFSLYLHAPCVITDP 360
DB 301 RMSWSLFLVHFGLSKRPENLAHSHVIFGPRYKLVNEIFNGRPLPDFSMYLSPCVITDP 360
QY 361 SLAPGCGSYVYLAPVPHLGTANLMDWAVEGPRLRDRIPDYLSCHYMPGLRSQILVTHRMFT 420
DB 361 SLAPGEGMSTHYVLAPVPHLGRADVWEAEPAGVAERIFEELERRAIPDRKHLTVSRIFS 420
QY 421 PEDFRDELNAWQGSAPSVPEILTQSAWEPHPNEDKHIDNLYLVGAGTHPGAGIPGVGSA 480
DB 421 PADFSTELSAHGSAPSVPEILTQSAWEPHPNEDRAIPNFIYVAGTHPGAGIPGVGSA 480
QY 481 KATAGLMEDL 491
DB 481 KATAQVMSDL 491

RESULT 9

US-09-298-718-5
Sequence 5, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-298-718-5

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEGFTFDAGPTVITDP 60

b 1 MSALVIGAGFGLALALRLQAGATTVIIEARDPGGRAYVWDDQGHVFDAGPTVWTD 60
61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWSGKVFYNDQAOLEAQIQOFPNPDVAG 120
61 DSIRELWALSGQPMERDVTLLPVSPFYRLTWADGRSFYVNDDBELIPQVASFNPADVDG 120
121 YRAFLDYGRAVENEGYKLGTVPELSFKDMLRAAPOLAKIQAWSVYSKVAGYIEDEHLR 180
121 YRRFDYAEVYREGYKLGTVPELSFKDMLRAAPOLAKIQAWSVYSKVAGYIEDEHLR 180
181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKILFQDLGGEVVL 240
181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKILFQDLGGEVVL 240
241 NARVHMETVGDKIQAQVLEDGRRETCAVASADVWHTYRDLLSQHPAAAKQAKKQSK 300
241 NARVTRIDTEGRATGVTLDDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAAILNRQ 300
301 RMSNSLFVLYFGLNHHQDLAHTVCFGRPYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
301 RMSNSLFVLYFGLNHHQDLAHTVCFGRPYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPLGRLSOLVTHRMFT 420
361 SLAPGCGMSTHYVLAAPVPHLGRADVDWEAEAPGYAERIFEELEERRAIPLDLRKHLTVSRI 420
421 PFDRELNAMQGSFVSVEPILTQSAMFRPNRKHIDNLYLVGNGTHPGAGIGVIGSA 480
421 PADFSTELSAHSGSAFVSVEPILTQSAMFRPNRKHIDNLYLVGNGTHPGAGIGVIGSA 480
481 KATAGLMLEDL 491
481 KATAQWMLSDL 491

RESULT 10

3-09-546-969-5
Sequence 5, Application US/09546969
Patent No. 6207409
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546,969
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-546-969-5

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

Qy 1 MKETTVIGAGFGLALALRLQAGATTVIIEARDPGGRAYVWDDQGHVFDAGPTVWTD 60
Db 1 MSALVIGAGFGLALALRLQAGATTVIIEARDPGGRAYVWDDQGHVFDAGPTVWTD 60
Qy 61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWSGKVFYNDQAOLEAQIQOFPNPDVAG 120
Db 61 DSIRELWALSGQPMERDVTLLPVSPFYRLTWADGRSFYVNDDBELIPQVASFNPADVDG 120
Qy 121 YRAFLDYGRAVENEGYKLGTVPELSFKDMLRAAPOLAKIQAWSVYSKVAGYIEDEHLR 180
Db 121 YRRFDYAEVYREGYKLGTVPELSFKDMLRAAPOLAKIQAWSVYSKVAGYIEDEHLR 180
Qy 181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKILFQDLGGEVVL 240
Db 181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKILFQDLGGEVVL 240
Qy 241 NARVHMETVGDKIQAQVLEDGRRETCAVASADVWHTYRDLLSQHPAAAKQAKKQSK 300
Db 241 NARVTRIDTEGRATGVTLDDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAAILNRQ 300
Qy 301 RMSNSLFVLYFGLNHHQDLAHTVCFGRPYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
Db 301 RMSNSLFVLYFGLNHHQDLAHTVCFGRPYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
Qy 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPLGRLSOLVTHRMFT 420
Db 361 SLAPGCGMSTHYVLAAPVPHLGRADVDWEAEAPGYAERIFEELEERRAIPLDLRKHLTVSRI 420
Qy 421 PFDRELNAMQGSFVSVEPILTQSAMFRPNRKHIDNLYLVGNGTHPGAGIGVIGSA 480
Db 421 PADFSTELSAHSGSAFVSVEPILTQSAMFRPNRKHIDNLYLVGNGTHPGAGIGVIGSA 480
Qy 481 KATAGLMLEDL 491
Db 481 KATAQWMLSDL 491

RESULT 11

US-08-980-832-4
Sequence 4, Application US/08980832B
Patent No. 6291204
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/08/980,832B
CURRENT FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-08-980-832-4

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGGGLALAIRLOAGIPVLLLEQDKPGGRAYVVOEGFTFDAGPTVITDP 60
1 MSSAIVIGAGGGLALAIRLOAGIATTIVEARDKPGGRAYVWDDQGHVDFDAGPTVITDP 60
61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFNTDQAOLEAQIQOQFNPRDVAG 120
61 DLSRELWALSGQPMERDVTLPPVSPFRLTWADGRSFYVNDDELIRQVASFNPADVDG 120
121 YRAPLDYSRAVFEVYEGYKLGTVFPLSKDMLRAAPOLAKLOAWRSVYSKVAGYIEDEHLR 180
121 YRPHDYAEVYREGYKLGTTFFLKGQMLNAPALMLRLOAYRSVHSMVARFIQDPHLR 180
181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFTLLVGGNPFSTSIYALIHALLERGGVWFPAKGGTNQLVAGWVALFELGGTLLL 240
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSOHPPAAKQAKLOSK 300
241 NARVTRIDTEGDRATGVTLLDGRQADTVASNGDVHVSYRDLGHTTRGRRTKAAILNRQ 300
301 RMSNSLFLVYFGLNHHHDLAHTVCGFPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RWSMSLFLVHFGLSKRPENLAHSEVIFGPYKGLVNEIFNGPELPDQFSWYLSHSPCVTDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDAVEGPRDLRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGMSTHYVLAPVPHLGRADVWEAEAPGVAERIFFELEERRAIPDLRKHLLTVSRIFS 420
421 PFDPRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLVYGAGTHPGAGTGPVIGSA 480
421 PADSTELSAHSGSAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGTGPVIGSA 480
481 KATAGLMLEDL 491
481 KATAQVMSDL 491

RESULT 12

US-09-547-267-5
Sequence 5, Application US/09547267
Patent No. 6613543
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: PERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-547-267-5
Query Match
Best Local Similarity 61.2%; Score 1595; DB 4; Length 494;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
QY 1 MKPTTVIGAGGGLALAIRLOAGIPVLLLEQDKPGGRAYVVOEGFTFDAGPTVITDP 60
DB 1 MSSAIVIGAGGGLALAIRLOAGIATTIVEARDKPGGRAYVWDDQGHVDFDAGPTVITDP 60
QY 61 SAIEELPALAGKOLKDYVELLPVTPYRLCWESGKVFNTDQAOLEAQIQOQFNPRDVAG 120
DB 61 DLSRELWALSGQPMERDVTLPPVSPFRLTWADGRSFYVNDDELIRQVASFNPADVDG 120
QY 121 YRAPLDYSRAVFEVYEGYKLGTVFPLSKDMLRAAPOLAKLOAWRSVYSKVAGYIEDEHLR 180
DB 121 YRPHDYAEVYREGYKLGTTFFLKGQMLNAPALMLRLOAYRSVHSMVARFIQDPHLR 180
QY 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFTLLVGGNPFSTSIYALIHALLERGGVWFPAKGGTNQLVAGWVALFELGGTLLL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSOHPPAAKQAKLOSK 300
DB 241 NARVTRIDTEGDRATGVTLLDGRQADTVASNGDVHVSYRDLGHTTRGRRTKAAILNRQ 300
QY 301 RMSNSLFLVYFGLNHHHDLAHTVCGFPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
DB 301 RWSMSLFLVHFGLSKRPENLAHSEVIFGPYKGLVNEIFNGPELPDQFSWYLSHSPCVTDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDAVEGPRDLRIFDYLEQHYMPGLRSQLVTHRMFT 420
DB 361 SLAPEGMSTHYVLAPVPHLGRADVWEAEAPGVAERIFFELEERRAIPDLRKHLLTVSRIFS 420
QY 421 PFDPRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLVYGAGTHPGAGTGPVIGSA 480
DB 421 PADSTELSAHSGSAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGTGPVIGSA 480
QY 481 KATAGLMLEDL 491
DB 481 KATAQVMSDL 491

RESULT 13

US-09-920-923B-4
Sequence 4, Application US/09920923B
Patent No. 6677134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920,923B
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patentin version 3.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923B-4
Query Match
Best Local Similarity 61.2%; Score 1595; DB 4; Length 494;

Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGGGLALALRLQAGIPVLLLEQKDPGGRAYVYQEGFTFDAGPTVITDP 60
1 MSSAIVIGAGGGLALALRLQAGIATIVEARDKPGGRAYVWMDQGHVFDAGPTVITDP 60
61 SAIBELPALAGKQKDYVELLPVTPFYRLCWESGVFNVDNDQALEAQIQQFNPRDVAG 120
61 DLSRELWALSGQPMERDVTLLPVSPFYELTWADGRSEYVNDDELIRQVASFNPADVQ 120
121 YRAFIDYSRVFNNEGILKGTVPFLSPKDMRLRAAPLAKLQAMRSVYKVGAYIEDEHLR 180
121 YRRFDYAEVYREGYLKGTTPFLKQMLNAAAPALMRLLQAVRSVHSMVARETQDPLR 180
181 QAFSPHSLVGVGNPATSSITLIHALEREGVWFPFGGTGALVNGMKLQDLSGEVVL 240
181 QAFSPHSLVGVGNPATSSITLIHALEREGVWFPFGGTGALVNGMKLQDLSGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRFETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
241 NARVTRIDTEGRATGVTLLDGRQLRADTVASNGDVMSYREDLIGHTRGRKTAAILNRQ 300
301 RMSNSLVLYGLNHHHDLAHTVCGPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLVLYGLNHHHDLAHTVCGPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPGCGSYVVLAPVPHLGTANLQWAVEGPRLDRIQVLEQHYMPGLRSQLVTHRMFT 420
361 SLAPGCGSYVVLAPVPHLGTANLQWAVEGPRLDRIQVLEQHYMPGLRSQLVTHRMFT 420
421 PPFDFELNAGWGSAPSVPEILTSQSAWFRPHNRDKHIDNLYVGAGTHPGAGIPGVIGSA 480
421 PADPSTELSAHGSAPSVPEILTSQSAWFRPHNRDKHIDNLYVGAGTHPGAGIPGVIGSA 480
481 KATAGLMEIDL 491
481 KATAQVMSDL 491

SU14
-09-934-903-16
Sequence 16, Application US/09934903
Patent No. 6660507
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odum, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: CL1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 16
LENGTH: 511
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF8
-09-934-903-16

Query Match 27.0%; Score 703.5; DB 4; Length 511;
Best Local Similarity 31.7%; Pred. No. 1.9e-61;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 2 KPTTVIGAGGGLALALRLQAGIPVLLLEQKDPGGRAYVYQEGFTFDAGPTVITDP 61
DB 5 KHIIVIGAGPGGLCAGMLLSORGFKVSIFDHAETGGRRNRPINMNGFTFDGPTFLMKG 64
QY 62 AIELFALAGKQKDYVELLPVTPFYRLCWESGVFNVDNDQALEAQIQQFNPRDVAG 121
DB 65 VLDEMFEELCERRSELYLEFLPLSPYRLLYDDRDIFVY-SDRENRAELQRFVDEGTDG 123
QY 122 RAPLDYSRAVENEGY-----LKLGTVPFLSPKDMRLRAAPLAKLQAMRSVYKVG 172
DB 124 EFMEOERRRNALYPCITRDYSSLK-----SFLSL-DLIKALPWLA---FPKSVFNMLGQ 175
QY 173 YIEDHLLQAFSPFLSLVGVGNPFATSSITLIHALEREGVWFPFGGTGALVNGMKLQ 232
DB 176 YFNOEKMLAFQCFQSKYLGMSWPCPALFTMLPYLEHEGYIYHVGKGLNRITAAAMAQVIA 235
QY 233 DLGGEVFNARVSEMETVGDKIQAQVLEDGRFETCAVASNADVVHTYRDLLSQHPAAAK 292
DB 236 ENGGEIHLNRSIESLIENGAAKGVKLGHAELRGDEVIINADFAHAMTHLVK--PGVLK 293
QY 293 Q-AKKIQSKMSNSLVLYGLNHHHDLAHTVCGPRYRELHIEIFNHDGLAEDFSL 350
DB 294 KYPENLQKREYSCTFMLYGLDKIYD-LPHHTIIVAKOYTTMIRNIFDNKTLTDDFSP 352
QY 351 YLHAPCVTDPSLAPGCGSYVVLAPVPHLGTANLQWAVEGPRLDRIQVLEQHY-MPGL 409
DB 353 YVQNASASDSDSLAPAGKALYLVMPN-NDSGLDQWQAHQCNVEQVLDLGLARLGLSDI 411
QY 410 RSOLVTHRMFTPEPFRDELNAGWGSAPSVPEILTSQSAWFRPHNRDKHIDNLYVGAGTHP 469
DB 412 RAHICEKLIIPQWETDEHYKATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHP 471
QY 470 GAGIPGVIGSAKATAGLM 487
DB 472 GSGLPVIESARISAKLI 489

RESULT 15
US-09-934-903-18
Sequence 18, Application US/09934903
Patent No. 6660507
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odum, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: CL1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 18
LENGTH: 497
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF9
US-09-934-903-18

Query Match 26.4%; Score 687; DB 4; Length 497;
Best Local Similarity 33.3%; Pred. No. 8e-60;
Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

QY 6 VIGAGGGLALALRLQAGIPVLLLEQKDPGGRAYVYQEGFTFDAGPTVITDPSAIEE 65
DB 11 VIGAGGGLSAISLATAGFSVQLIKRNDKVGKLNIMTKDGTFDLQPSILTMPHFEA 70

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:26:38 ; Search time 58.1706 Seconds
(without alignments)
2389.754 Million cell updates/sec

File: US-09-941-947A-32

Effect score: 2606
Sequence: 1 MKPTTVIGAGFGGLAIRL.....IPGVIGSAKATAGLMLEDLI 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseq 29Jan04:*
- 2: Geneseq1980s:*
- 3: Geneseq1990s:*
- 4: Geneseq2000s:*
- 5: Geneseq2001s:*
- 6: Geneseq2002s:*
- 7: Geneseq2003as:*
- 8: Geneseq2003bs:*
- 9: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	100.0	492	5	AEE22314
2	2606	100.0	492	6	AAO16021
3	2606	100.0	492	6	ABP96688
4	2438	93.6	492	2	AAE07466
5	2438	93.6	492	2	AAW82257
6	2438	93.6	492	2	AAW87889
7	2438	93.6	492	2	AAW99099
8	2438	93.6	492	2	AAW26333
9	2050	78.7	494	6	ABW70123
10	1849.5	71.0	489	2	AAW01122
11	1849.5	71.0	489	2	AAW00173
12	1849.5	71.0	489	2	AAW32472
13	1845.5	70.8	489	2	AAW13985
14	1845.5	70.8	489	2	AAW13984
15	1845.5	70.8	489	2	AAW01123
16	1845.5	70.8	489	2	AAW00174
17	1845.5	70.8	489	2	AAW32475
18	1595	61.2	494	2	AAW00871
19	1592	61.1	494	2	AAW69332
20	1415	54.3	526	2	AAW95897
21	703.5	27.0	511	5	AAE22309
22	703.5	27.0	511	5	ABG61588
23	703.5	27.0	511	5	ABG61588
24	703.5	27.0	511	6	ADA14534
25	687	26.4	497	5	AAE22310

26	687	26.4	497	5	ABG61589	ABG61589	High grow
27	687	26.4	497	5	AAU80333	AAU80333	Methylomo
28	687	26.4	497	5	ADA14538	ADA14538	Methylomo
29	624	23.9	502	6	ABW72613	ABW72613	Staphyloc
30	624	23.9	502	6	ADA14542	ADA14542	Staphyloc
31	598.5	23.0	582	6	ABP97465	ABP97465	Blakeslea
32	592	22.7	499	2	AAE64269	AAE64269	Anabaena
33	587	21.8	490	5	ABW49224	ABW49224	Listeria
34	567	21.8	490	6	ABU32531	ABU32531	Protein e
35	562	21.6	582	2	AAW22499	AAW22499	Phaffia d
36	542	20.8	502	6	ABU43877	ABU43877	Protein e
37	531	20.4	544	5	AAO15518	AAO15518	Agromyces
38	492	18.9	543	5	AAO15521	AAO15521	Micrococc
39	484	18.6	497	6	ADA14544	ADA14544	Staphyloc
40	484	18.6	500	6	ABW72616	ABW72616	Staphyloc
41	482	18.5	497	6	ABU16031	ABU16031	Protein e
42	461.5	17.7	530	6	AAE31690	AAE31690	Rhodococc
43	416.5	16.0	548	4	AAW85728	AAW85728	Enzyme in
44	415	15.9	439	4	AAU34200	AAU34200	Staphyloc
45	410.5	15.8	548	4	AAW76640	AAW76640	Corynebac

ALIGNMENTS

RESULT 1
AAE22314
ID AAE22314 standard; protein; 492 AA.

XX AAE22314;

XX (first entry)

DE Pantoea stewartii phytoene desaturase (CrtL) enzyme.

KW Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;
KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KW aquaculture; enzyme; phytoene desaturase; CrtL.

CS Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-C229859P.

XX 01-SEP-2000; 2000US-C229907P.

(DUPO) DU PONT DE NEMOURS & CO E I.

PI Brzostowicz PC, Cheng O, Dicosimo DJ, Koffas M, Miller ES;

PI Odum JM, Picataggio SK, Rouviere PE,

XX MPI; 2002-351711/38.

XX N-PSDB; AAD35512.

PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

PS Claim 19; Page 141-143; 156pp; English.

CC The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the control of regulatory sequences, and contacting the host cell with carbon substrate to produce a carotenoid compound. The method is useful for producing carotenoid compounds such as anthraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in

C the carotenoid biosynthetic pathway and which metabolise single carbon
C substrates. The carotenoids have potent anti-oxidant properties useful in
C diet, and aquaculture elements. The carotenoids are also useful as
C intermediates in the synthesis of steroids flavours and fragrances and
C compounds for potential electro-optic applications. The present sequence
C is Pantoea stewartii phytoene desaturase (CrtL) enzyme used in the
C invention

X Q Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.4e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGTVIDP 60
1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGTVIDP 60
61 SAIEELPALAGKQKDYVELLPVTPFVRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
61 SAIEELPALAGKQKDYVELLPVTPFVRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
121 YRAPLDYSRAVFNESGKYLKGTVPFLSFKDMRLAAPOLAKLQAWRSYVSKVAGVIEDEHLR 180
121 YRAPLDYSRAVFNESGKYLKGTVPFLSFKDMRLAAPOLAKLQAWRSYVSKVAGVIEDEHLR 180
181 QAFSFHSLLVGGNPFATSSYITLIHALEREWGVVFPFGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFHSLLVGGNPFATSSYITLIHALEREWGVVFPFGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAAKQKQSK 300
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAAKQKQSK 300
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
421 PFDFRDELNAWQGSFAFSEVILTQSAMFRPHNDKIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSFAFSEVILTQSAMFRPHNDKIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 2
AA016021
ID AA016021 standard; protein; 492 AA.

AC AA016021;

DT 20-FEB-2003 (first entry)

Pantoea stewartii phytoene desaturase.

Carotenoid; crt.

Pantoea stewartii.

PN WO200279395-A2.

PD 10-OCT-2002.

PF 25-JAN-2002; 2002WO-US002124.

PR 26-JAN-2001; 2001US-0264329P.

PR 04-MAY-2001; 2001US-0288984P.

XX (CRGI) CARGILL INC.
XX PA
XX De Souza ML, Kollmann SR, May CA, Schroeder WA;
XX WP; 2003-075455/07.
XX DR N-PSDB; ABT14193.
XX

Novel isolated nucleic acid useful e.g. to engineer host cells with the
PT ability to produce particular carotenoids and polypeptides useful in cell
PT -free systems to make particular carotenoids.

PS Claim 26; Page 63-64; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
CC carotenoid (crt)-related proteins. The crt-related DNA and protein
CC sequences of the invention are useful for engineering cells which are
CC able to produce carotenoids. The present amino acid sequence represents a
CC crt-related protein of the invention

XX Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 6; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.4e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGTVIDP 60
DB 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGTVIDP 60
QY 61 SAIEELPALAGKQKDYVELLPVTPFVRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
DB 61 SAIEELPALAGKQKDYVELLPVTPFVRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
QY 121 YRAPLDYSRAVFNESGKYLKGTVPFLSFKDMRLAAPOLAKLQAWRSYVSKVAGVIEDEHLR 180
DB 121 YRAPLDYSRAVFNESGKYLKGTVPFLSFKDMRLAAPOLAKLQAWRSYVSKVAGVIEDEHLR 180
QY 181 QAFSFHSLLVGGNPFATSSYITLIHALEREWGVVFPFGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFHSLLVGGNPFATSSYITLIHALEREWGVVFPFGGTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAAKQKQSK 300
DB 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAAKQKQSK 300
QY 301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNHDGLAEDFSLYLHAPCVTDP 360
DB 301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNHDGLAEDFSLYLHAPCVTDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
QY 421 PFDFRDELNAWQGSFAFSEVILTQSAMFRPHNDKIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDELNAWQGSFAFSEVILTQSAMFRPHNDKIDNLYLVGAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDLI 492
DB 481 KATAGLMLEDLI 492

RESULT 3
ABP96688
ID ABP96688 standard; protein; 492 AA.

XX ABP96688;

AC ABP96688;

DT 03-JUN-2003 (first entry)

Pantoea stewartii phytoene desaturase SEQ ID NO:8.

Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
crtI; crtB; crtZ; phytoene desaturase; enzyme; phytoene; carotenoid.
Pantoea stewartii.

WO2003016503-A2.

27-FEB-2003.

15-AUG-2002; 2002WO-US026647.

15-AUG-2001; 2001US-0312646P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

WPI; 2003-268323/26.

N-PSDB; ACC44762.

Novel nucleic acid molecule isolated from *Pantoea stewartii* encoding a carotenoid biosynthetic enzyme, useful for regulating carotenoid biosynthesis in an organism.

Claim 4; Page 63-65; 60pp; English.

The present invention describes *Pantoea stewartii* carotenoid biosynthetic enzymes (I). More specifically described are the geranylgeranyl pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX), lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase (crtS) and beta-carotene hydroxylase (crtZ) enzymes (I). (See ASP96685 to ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating carotenoid biosynthesis in an organism, by over-expressing (I) in an organism, such that the carotenoid biosynthesis is altered in the organism. (I) and the genes encoding (I) are useful for converting phytoene to the carotenoids, for creating recombinant organisms that have the ability to produce various carotenoid compounds, and also for enhancing or manipulating carotenoid compounds. (I) can also be used for producing gene products having enhanced or altered activity

Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 6; Length 492;

Best Local Similarity 100.0%; Pred. No. 4.4e-254;

Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTIVTDP 60

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTIVTDP 60

61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

121 YRAFLDYSAVNEGVYKLGTVTFPLSFYDMLRPAAPQLAKLQAWRSVYKVGATYEDHLR 180

121 YRAFLDYSAVNEGVYKLGTVTFPLSFYDMLRPAAPQLAKLQAWRSVYKVGATYEDHLR 180

181 QAFSFSLLVGNPNFATSSIYTLIHALLREMGVWPPRGCTGALVNGMIKLFQDLGGVWL 240

181 QAFSFSLLVGNPNFATSSIYTLIHALLREMGVWPPRGCTGALVNGMIKLFQDLGGVWL 240

241 NARVSHMETVGDKIQAQVLEDCRRRETCAVASNADVHTYRDLLSQHPAAAKQAKLQSK 300

241 NARVSHMETVGDKIQAQVLEDCRRRETCAVASNADVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFLVYFGLNHHHDLAHTTVCFGRYRELTHIEIFNHDGLAEDPSLYLHAPCVTDP 360

301 RMSNSLFLVYFGLNHHHDLAHTTVCFGRYRELTHIEIFNHDGLAEDPSLYLHAPCVTDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLQNAVEGPRRLDRIDYLEQHYMPGLRSQLYTHRMFT 420

361 SLAPEGCGSYVVLAPVPHLGTANLQNAVEGPRRLDRIDYLEQHYMPGLRSQLYTHRMFT 420

QY 421 PFDPRDELNAWQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDPRDELNAWQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

QY 481 KATAGLMLEDLI 492

DB 481 KATAGLMLEDLI 492

RESULT 4

AA07466
ID AA07466 standard; protein; 492 AA.

XX AA07466;

XX 24-OCT-2003 (revised)

DT 28-JAN-1991 (first entry)

XX Polypeptide with enzymatic activity for the conversion of phytoene into lycopene.

XX Carotenoid biosynthesis; vitamin A; cancer; food coloring.

XX *Pantoea ananatis*.

PN BP393690-A.

PD 24-OCT-1990.

PF 20-APR-1990; 90EP-00107493.

XX 21-APR-1989; 89JP-00103078.

PR 05-MAR-1990; 90JP-00053255.

XX (KIRI) KIRIN BEER KK.

XX Misawa N, Kobayashi K, Nakamura K;

XX WPI; 1990-32212/43.

DR N-PSDB; AAQ06296.

XX DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

PS Claim 4; Fig 4; 40pp; English.

XX Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;

Best Local Similarity 93.5%; Pred. No. 4.3e-237;

Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTIVTDP 60

DB 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTIVTDP 60

QY 61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

DB 61 SAIEELFALAGKQLKDYVELLPVTFFYRLCWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

QY 121 YRAFLDYSAVNEGVYKLGTVTFPLSFYDMLRPAAPQLAKLQAWRSVYKVGATYEDHLR 180

DB 121 YRAFLDYSAVNEGVYKLGTVTFPLSFYDMLRPAAPQLAKLQAWRSVYKVGATYEDHLR 180

QY 181 QAFSFSLLVGNPNFATSSIYTLIHALLREMGVWPPRGCTGALVNGMIKLFQDLGGVWL 240

DB 181 QAFSFSLLVGNPNFATSSIYTLIHALLREMGVWPPRGCTGALVNGMIKLFQDLGGVWL 240

241 NARVSHMETVGDKIQAQVLEDEGRRFETCAVASNADVVHTYRDLISQHPAAQAOKLQSK 300
 241 NARVSHMETVGNKIEAVHLEDEGRRFETQAVASNADVVHTYRDLISQHPAAVQSKLQTK 300
 301 RMSNSLVLYFGLNHHDDQLAHHTVCGPRYRELIIHFIHNDGLAEDFSLYLHAPCVTDP 360
 301 RMSNSLVLYFGLNHHDDQLAHHTVCGPRYRELIIHFIHNDGLAEDFSLYLHAPCVTDS 360
 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLSCHYMPGLRSQLVTHRMFT 420
 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLSCHYMPGLRSQLVTHRMFT 420
 421 PFDFRDLNNAWQGSASFSEVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 421 PFDFRDLNNAWQGSASFSEVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 481 KATAGLMLEDLI 492
 481 KATAGLMLEDLI 492
 RESULT 5
 AAW82257
 D AAW82257 standard; protein; 492 AA.
 X C AAW82257;
 X X
 T 17-OCT-2003 (revised)
 T 16-JUL-1999 (first entry)
 X X
 X C. utilis crtI protein.
 X X
 W HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtI;
 W carotenoid.
 X Pichia jadinii.
 X JPI0248575-A.
 X X
 D 22-SEP-1998.
 X X
 F 12-MAR-1997; 97JP-00058012.
 R 12-MAR-1997; 97JP-00058012.
 X (KIRI) KIRIN BREWERY KK.
 X WPI; 1998-560727/48.
 X N-PSDB; AAV73181.
 X Gene useful for increase in carotenoid production - and preparation of
 X carotenoid.
 X Example 2; Fig 11-14; 54pp; Japanese.
 X
 X This invention describes a novel method for the preparation of
 X carotenoids using genes and proteins isolated from *Candida utilis*. The
 X invention specifically describes the isolation of a 3-hydroxy-3-
 X methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence
 X represents the *Candida utilis* crtI protein which is used in the method of
 X the invention. (Updated on 17-OCT-2003 to standardise OS field)
 X Sequence 492 AA;
 X
 X Query Match 93.6%; Score 2438; DB 2; Length 492;
 X Best Local Similarity 93.5%; Pred. No. 4.3e-237;
 X Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;
 X
 Y 1 MKPTTVIGAGFGGALAIRQAAGIPVLLLEQKPKGGRAYVYCEQGFTEAGPTVITDP 60
 Y 1 MKPTTVIGAGFGGALAIRQAAGIPVLLLEQKPKGGRAYVYCEQGFTEAGPTVITDP 60
 Y 61 SAIEELFALAGKQKDYVELLPVTFFYRLCWESGKVENYDNDQTRLEAQIQQFNPRDVAG 120

DB 61 SAIEELFALAGKQKDYVELLPVTFFYRLCWESGKVENYDNDQTRLEAQIQQFNPRDVAG 120
 QY 121 YRAFLDYSRVAFNEGVLKLTVPFSLFKDMLRAAPOLAKLQAKRSVYSKVAGYIEDEHLR 180
 DB 121 YROFLDYSRVAFNEGVLKLTVPFSLFKDMLRAAPOLAKLQAKRSVYSKVAGYIEDEHLR 180
 QY 181 QAFSFFHLLVGGNPFATSSITTLIHALEREWGVPFRGGTGALVNGMIKLFQDLGGEVVL 240
 DB 181 QAFSFFHLLVGGNPFATSSITTLIHALEREWGVPFRGGTGALVNGMIKLFQDLGGEVVL 240
 QY 241 NARVSHMETVGDKIQAQVLEDEGRRFETCAVASNADVVHTYRDLISQHPAAQAOKLQSK 300
 DB 241 NARVSHMETVGNKIEAVHLEDEGRRFETQAVASNADVVHTYRDLISQHPAAVQSKLQTK 300
 QY 301 RMSNSLVLYFGLNHHDDQLAHHTVCGPRYRELIIHFIHNDGLAEDFSLYLHAPCVTDP 360
 DB 301 RMSNSLVLYFGLNHHDDQLAHHTVCGPRYRELIIHFIHNDGLAEDFSLYLHAPCVTDS 360
 QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLSCHYMPGLRSQLVTHRMFT 420
 DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLSCHYMPGLRSQLVTHRMFT 420
 QY 421 PFDFRDLNNAWQGSASFSEVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 DB 421 PFDFRDLNNAWQGSASFSEVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 QY 481 KATAGLMLEDLI 492
 DB 481 KATAGLMLEDLI 492
 RESULT 6
 AAW87889
 ID AAW87889 standard; protein; 492 AA.
 X X
 AC AAW87889;
 X X
 DT 17-OCT-2003 (revised)
 DT 10-MAR-1999 (first entry)
 X X
 DE Protein encoded by the carotenoid biosynthesis gene crtI.
 X X
 X Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;
 X crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;
 X carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
 X food additive.
 X Pantoea ananatis.
 X JPI0327865-A.
 X 15-DEC-1998.
 X 29-MAY-1997; 97JP-00140460.
 X 29-MAY-1997; 97JP-00140460.
 X (KIRI) KIRIN BREWERY KK.
 X (KAIY-) XAIYO BIOTECHNOLOGY KENKYUSHO KK.
 X WPI; 1999-099030/09.
 X N-PSDB; AAV84081.
 X New carotenoid glucoside(s) - used as food additives.
 X Disclosure; Page 18-20; 26pp; Japanese.
 X
 X The present sequence represents a protein involved in carotenoid
 X biosynthesis. The specification describes astaxanthin diglucosides and
 X adonixanthin-3'-glucosides. The specification also describes a method for
 X the preparation of a carotenoid glycoside, in which all, or part of,
 X carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW

are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;
Best Local Similarity 93.5%; Pred. No. 4.3e-237;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;
1 MKPTTVIAGGFGGLALAIRLOAGIPVLLLEQORDKPGRAYVYQEGTFDAGPTVITDP 60
1 MKPTTVIAGGFGGLALAIRLOAGIPVLLLEQORDKPGRAYVYQEGTFDAGPTVITDP 60
61 SAIEELFALAGKQKLDYVELLPVTPFYRLCWESKVFYNDNDQALBAQIQQFNPRDVAG 120
61 SAIEELFALAGKQKLDYVELLPVTPFYRLCWESKVFYNDNDQALBAQIQQFNPRDVAG 120
121 YRAFIDYGRAVFNEGYLKLGTVPFSLPKDMRLAAPOLAKLQAMRSVYSKVAGYIEDEHLR 180
121 YRQFLDYGRAVFEKGYLKLGTVPFLSPFDMRLAAPOLAKLQAMRSVYSKVAGYIEDEHLR 180
181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMIKLPQDLGGSEVL 240
181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMIKLPQDLGGSEVL 240
241 NARVSHMETVGDGKIQAQVLEDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
241 NARVSHMETTGNKIEAVHLEDGRFRFLTQAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELHIEIENHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELHIEIENHDGLAEDFSLYLHAPCVTDS 360
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPKLRDRIPAYILEQHYMPGLRSQLVTHRMFT 420
421 PFDRELNAMQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDRELNAMQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 7

AAW99099 standard; protein; 492 AA.
AAW99099;
17-OCT-2003 (revised)
14-MAY-1999 (first entry)

Erwinia uredovora crtI protein sequence.

Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;

metabolite.

Pantoea ananatis.

JPL1046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

07-AUG-1997; 97JP-00213648.

(KIRI) KIRIN BREWERY KK.

WPI; 1999-208113/18.

N-PSDB; AAX19119.

Beta-carotin hydroxylase - useful for preparation of xanthophylls and their metabolites.

Disclosure; Page 13-15; 17pp; Japanese.

The present invention describes beta-carotin hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crtI protein sequence from the present invention. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;
Best Local Similarity 93.5%; Pred. No. 4.3e-237;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;
1 MKPTTVIAGGFGGLALAIRLOAGIPVLLLEQORDKPGRAYVYQEGTFDAGPTVITDP 60
1 MKPTTVIAGGFGGLALAIRLOAGIPVLLLEQORDKPGRAYVYQEGTFDAGPTVITDP 60
61 SAIEELFALAGKQKLDYVELLPVTPFYRLCWESKVFYNDNDQALBAQIQQFNPRDVAG 120
61 SAIEELFALAGKQKLDYVELLPVTPFYRLCWESKVFYNDNDQALBAQIQQFNPRDVAG 120
121 YRAFIDYGRAVFNEGYLKLGTVPFSLPKDMRLAAPOLAKLQAMRSVYSKVAGYIEDEHLR 180
121 YRQFLDYGRAVFEKGYLKLGTVPFLSPFDMRLAAPOLAKLQAMRSVYSKVAGYIEDEHLR 180
181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMIKLPQDLGGSEVL 240
181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMIKLPQDLGGSEVL 240
241 NARVSHMETVGDGKIQAQVLEDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
241 NARVSHMETTGNKIEAVHLEDGRFRFLTQAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELHIEIENHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELHIEIENHDGLAEDFSLYLHAPCVTDS 360
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPKLRDRIPAYILEQHYMPGLRSQLVTHRMFT 420
421 PFDRELNAMQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDRELNAMQGSFAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 8

AAW26333 standard; protein; 492 AA.
AAW26333;

17-OCT-2003 (revised)

13-JAN-2000 (first entry)

Erwinia uredovora phytoene desaturase enzyme (PDS).

Transgenic plant; herbicide; resistant; tolerant; phytoene desaturase;

PDS; crtI gene; 4-hydroxy-phenylpyruvate dioxygenase; 4HPPD; banana;

cotton; maize; tomato; vine; weed control; isoxazole herbicide family;

N-PSDB; AAT91544.

DNA encoding Erwinia herbicola lycopene cyclase - for producing recombinant enzyme, and transgenic organisms with increased beta-carotene levels.

Example 10; Fig 11; 102pp; English.

A novel DNA molecule has been isolated which encodes an Erwinia herbicola lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA molecule comprises at least 1125 bp and is present in the plasmids pAR147, pAR1509, pAR1510 and pAR1520. The present sequence represents the amino acid sequence corresponding to the structural gene for phytoene dehydrogenase-4H. The new DNA molecule can be used to produce the recombinant enzyme and transgenic organisms, e.g. yeasts or plants, with increased beta-carotene levels. Beta-carotene is used as a colourant in margarine and butter and as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to standardise OS field)

Sequence 489 AA;

Query Match 71.0%; Score 1849.5; DB 2; Length 489;
Best Local Similarity 72.3%; Pred. No. 1.4e-177;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLAALRLQAGIPVLLLEQDQKPGGRAYVYQGGTFDAGTVIDP 60
1 MKPTTVIGAGFGGLAALRLQAGIPVLLLEQDQKPGGRAYVYQGGTFDAGTVIDP 60

61 SAIEELFALAGKQLDYVLLPVTFFYRLCWESGKVFVNDQAOLEAQIOQFNPRDVAG 120
61 TALEALFTLAGRMEDYVLLPVKPYRLCWESGKTLDYANDSFLEAQITQFNPRDVAG 120

121 YRAFIDYSRAVFNEGYKLGTVPFLLSFKMDLRAAPQALQAWRSVYSKVAGYIEDEHLR 180
121 YRRFLAYSQAVFQGYLRGSPVFLSFRDMLRAGPQLLQAWQSVYQSVSRFIEDEHLR 180

181 QAFSFSHLLVGGNPFATSSIYTLIHALERWGVWPFPGGTGALVNGMVKLFQDLGGEVVL 240
181 QAFSFSHLLVGGNPFATSSIYTLIHALERWGVWPFPGGTGALVNGMVKLFQDLGGEVVL 240

241 NAFVSHMETVGDKIQAQVLEDDGRPFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300
241 NAFVSHMETVGDKIQAQVLEDDGRPFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFLVYFGLNHHDDQLAHHTVTCFGRPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFLVYFGLNHHDDQLAHHTVTCFGRPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360

361 SLAPEGCGSVYVLAIVPHLGTANLQVAVGPRRLRDRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSVYVLAIVPHLGTANLQVAVGPRRLRDRIFDYLEQHYMPGLRSQLVTHRMFT 420

421 PFDRELNAWQGSFVSPEIITQSNAWPHRNRKHIQNLVVGAGTHPGAGIPGVIGSA 480
421 PFDRELNAWQGSFVSPEIITQSNAWPHRNRKHIQNLVVGAGTHPGAGIPGVIGSA 480

481 KATAGLWLSL 491
478 ESTASLWLSL 488

RESULT 13
AR13985
D AR13985 standard; protein; 489 AA.
X C AAR13985;
X T 24-OCT-2003 (revised)
T T 25-MAR-2003 (revised)
T T 26-NOV-1991 (first entry)

Phytoene dehydrogenase-4H from pARC146D.

GGPP; carotenoid; phytoene; zeaxanthin; lycopene.

Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).

WO9113078-A.

05-SEP-1991.

04-MAR-1991; 91WO-US001458.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

(STAD) AMOCO CORP.

Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
Yen HC;

WPI; 1991-281410/38.

N-PSDB; AAQ13720.

Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA encoding enzymes from Erwinia herbicola.

Disclosure; Fig 15(1-4); 313pp; English.

There are a total of six relevant genes in a 7900 bp region that cause E. coli cells to produce GGPP and the carotenoids phytoene through zeaxanthin diglucoside, which is the final prod. identified in the carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607 (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase, phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase are represented in: AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726 respectively. Recombinant expression plasmids can be used to produce large amounts of the enzymes and hence large amounts of the carotenoids which they synthesise. Carotenoids are pigments with a variety of applications. CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 24-OCT-2003 to standardise OS field)

Sequence 489 AA;

Query Match 70.8%; Score 1845.5; DB 2; Length 489;
Best Local Similarity 72.1%; Pred. No. 3.5e-177;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLAALRLQAGIPVLLLEQDQKPGGRAYVYQGGTFDAGTVIDP 60
1 MKPTTVIGAGFGGLAALRLQAGIPVLLLEQDQKPGGRAYVYQGGTFDAGTVIDP 60

61 SAIEELFALAGKQLDYVLLPVTFFYRLCWESGKVFVNDQAOLEAQIOQFNPRDVAG 120
61 TALEALFTLAGRMEDYVLLPVKPYRLCWESGKTLDYANDSFLEAQITQFNPRDVAG 120

121 YRAFIDYSRAVFNEGYKLGTVPFLLSFKMDLRAAPQALQAWRSVYSKVAGYIEDEHLR 180
121 YRRFLAYSQAVFQGYLRGSPVFLSFRDMLRAGPQLLQAWQSVYQSVSRFIEDEHLR 180

181 QAFSFSHLLVGGNPFATSSIYTLIHALERWGVWPFPGGTGALVNGMVKLFQDLGGEVVL 240
181 QAFSFSHLLVGGNPFATSSIYTLIHALERWGVWPFPGGTGALVNGMVKLFQDLGGEVVL 240

241 NAFVSHMETVGDKIQAQVLEDDGRPFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300
241 NAFVSHMETVGDKIQAQVLEDDGRPFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFLVYFGLNHHDDQLAHHTVTCFGRPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFLVYFGLNHHDDQLAHHTVTCFGRPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360

```

301 SMSNSLVFLVFGNLQNPHSLAHHTTICGPRPYRELIDEIFTGSALADDFSLYLHSPCVTDP 360
361 SLAPEGCGSYVTLAPVPHLGTANLDMAVEGPRRLDRIFDYLEGHYMPGLRSQLVATRMFT 420
361 SLAPPCCASFVTLAPVPHLGNAPLDMAQEGPKLRDRIFDYLEERYMPEGLRSQLVATQRIFT 420
421 PFDFERDELNAWQSAPSFVEPILTQSAWFRPHNRDKHIDNLVLYNCAGTHPGAGIPGVIGSA 480
421 RQTSRHWAIILGSLFTEPPSLTQGLF--AANATRH-SNDLYVAAGTHPGAGIPGVVGLA 477
481 KATAGLMLEDL 491
478 ESTASLMIEDL 488

RESULT 14
ARI13984
D AAR13984 standard; protein; 489 AA.
X X
C C AAR13984;
T T 24-OCT-2003 (revised)
T T 25-MAR-2003 (revised)
T T 26-NOV-1991 (first entry)
X X
E E Phytoene dehydrogenase-4H.
X X
W W GAPP; carotenoid; phytoene; zeaxanthin; lycopene.
X X
S S Pantosa agglomerans; EHO-10 (E. vulneris - ATCC 39368).
X X
N N WO9113078-A.
X X
D D 05-SEP-1991.
X X
E E 04-MAR-1991; 91WO-US001458.
X X
R R 02-MAR-1990; 90US-00487613.
R R 18-MAY-1990; 90US-00525551.
R R 03-AUG-1990; 90US-00562674.
R R 28-FEB-1991; 91US-00662921.
X X
X X (STAD ) AMOCO CORP.
X X
Y Y Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
Y Y Yen HC;
X X
X X WPI; 1991-28:410/38.
R R N-PSDB; AAQ13719.

```

Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA encoding enzymes from *Erwinia herbicola*.

Disclosures: Fig 11(1-4): 313pp; English.

There are a total of six relevant genes in a 7900 bp region that cause E. coli cells to produce GGPP and the carotenoids phytylene through zeaxanthin diglycoside, which is the final prod. identified in the zeaxanthin diglycoside contd. in plasmid PARC376 (contg. a ca. 13 kb chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607 (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase, phytylene synthase, phytylene dehydrogenase-4H, lycopene cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase are represented in AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726 respectively. Recombinant expression plasmids can be used to produce large amts. of the enzymes and hence large amts. of the carotenoids which they synthesize. Carotenoids are pigments with a variety of applications. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 24-OCT-2003 to standardise QS field)

Sequence 489 AA:

Query Match: 70.88; Score 1845.5; D3 2: Length 489;

Best Local Similarity	72.1%;	Pred. No..3.5e-177;	
Matches	354;	Conservative	45; Mismatches 89; Indels 3; Gaps 2

QY	1	MKPTTVI	GAGGGGALAIRLQAAGIPVLLLEQSDKPGGRAYVYQSGFTFDAGTVIDTP	60
DB	1	MKTTVI	GAGGGGALAIRLQAAGIPTVLLLEQSDKPGGRAYVYHDDGQFTFDAGTVIDTP	60
QY	61	SAIRELP	FALQKQKQKDYVELLPVTFVRLCWSSGKVFYNDQAOLEAQIQCFNPRDVAG	120
DB	61	TALEALF	FLAGRRMDEVRLVPKPFYRLCWSSGKTLTDANDSFELEAQITQFNPRDVAG	120
QY	121	YRAFIDY	GRAVFNENGYLKLGTVPFLSFKMLRAAPQAKLQAWRSYYSKVAGYIEDEHLR	180
DB	121	YERFLAY	QAVFQGYLRLGSPFLSFDKMLRAGPQLKLQAWQSYYSQSVSRFIDEHLR	180
QY	181	QAPSPHSL	LVGNPPATSSITLIHALERENGWFRPGGTGALVNGMKLFODLGGEVVL	240
DB	181	QAPSPHSL	LVGNPFYSSITLIHALERENGWFPFGGTGALVNGMKLFTDLGGEL	240
QY	241	NARVSHMETV	GDKTQIAQVLEDGRGFETCAVASNADVVTYTRDLLSHQPAAKQAQKQASK	300
DB	241	NARVEELV	ADNRVRSQVRLADGRIFUTDAVASNADVVTYTKLLGTIPVQKRAARLERK	300
QY	301	RMNSNLF	VLVTGLNHHDDQLAHTTFCGPRYRRLIHFIHNDGLAEDFSLYLHAPCVTDP	360
DB	301	SMNSNLF	VLVTGLNQPHSQLAHTTFCGPRYRRLIDEIFTGSALADDFSLYLHSPCVTDP	360
QY	361	SIAPGCGSY	VYLAAPVPELGTANLDNAVCGPRLRDRIFDYLBQHYMPGLRSOLVTHRMFT	420
DB	361	SIAPPCCAS	FYLAAPVPELGNAPLDNAQSGFKURDRIFDYLBQHYMPGLRSOLVTHRMFT	420
QY	421	PRDFRDEL	NWQGSAFSVEPILTQSAWFRPHNRKHDNLVYLVGATHPGAGIPGVIGSA	480
DB	421	ROTSRHA	NIATLGSILFTPPSLTQGLF--AANATRH-SNLVYLVGATHPGAGIPGVVGLA	477
QY	481	KATAGIM	LEDL 491	
DB	478	ESTASIM	LEDL 488	

RESULT	15
AAW01123	
ID	AAW01123 standard; protein; 489 AA.
XX	AAW01123;
AC	
XX	
XX	
DT	16-OCT-2003 (revised)
DT	25-MAR-2003 (revised)
DT	11-DEC-1996 (first entry)
XX	
XX	
DE	Phytoene dehydrogenase-4H encoded on pARC1460.
XX	
XX	
KW	GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
KW	phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
KW	pigment; food colourant; chloroplast transit peptide; increase yield;
KW	tobacco ribulose bis-phosphate carboxylase-oxygenase.
XX	
OS	Pantoea agglomerans.
XX	
XX	US5530188-A.
PN	
XX	
PD	25-JUN-1996.
XX	
PF	21-JUL-1993; 93US-00095726.
XX	
XX	
PR	02-MAR-1990; 90US-00487613.
PR	18-MAY-1990; 90US-00525551.
PR	03-AUG-1990; 90US-00562674.
PR	28-FEB-1991; 91US-00562921.
PR	30-OCT-1991; 91US-00785566.
XX	
XX	(STAD) AMOCO CORP.
PA	
XX	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 31.4106 Seconds
(without alignments)
3837.172 Million cell updates/sec

file: US-09-941-947a-30

effect score: 2021

quence: 1 MDPHYDLILVAGLGLIA.....SGRPPVPFAALQAIMTTR 382

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518282 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mbc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	ID	Description
1	2021	100.0	382	2	Q8GCS2
2	1715	84.9	382	2	Q47844
3	1494	73.9	386	2	Q8VUJ6
4	899	44.5	385	2	Q93C18
5	804	39.8	394	2	Q9KIX3
6	783.5	38.8	382	2	P94791
7	735.5	36.4	386	2	Q9RLH5
8	637	31.5	434	2	Q06756
9	156.5	7.7	498	10	Q9M424
10	152.5	7.5	498	10	Q9FV32
11	150	7.4	524	10	Q9XGX3
12	149	7.4	490	10	Q84V99
13	149	7.4	504	10	Q9M546
14	148.5	7.3	418	2	P72451
15	146	7.2	504	10	Q8LPP7
16	146	7.2	504	10	Q8GTR2

17	145.5	7.2	498	10	Q9LWA6
18	144.5	7.1	417	16	Q7U994
19	142.5	7.1	502	10	Q9AXL1
20	139.5	6.9	430	10	Q8L8H5
21	139.5	6.9	511	10	Q9FV42
22	138.5	6.9	495	10	Q8S3C3
23	138	6.8	902	2	Q8KIP3
24	137.5	6.8	437	10	Q8S3J4
25	137	6.8	377	2	Q7WT70
26	136.5	6.8	517	10	Q8VWR6
27	136	6.7	540	10	Q8LJ81
28	134.5	6.7	516	10	Q9FV43
29	133	6.6	407	16	Q88C14
30	130.5	6.5	529	10	Q9AXL0
31	130.5	6.5	594	10	Q7XAV8
32	130	6.4	549	16	Q8PCT2
33	129.5	6.4	529	10	Q9AXK9
34	128	6.3	399	16	Q7V6N0
35	127.5	6.3	399	16	O06934
36	127.5	6.3	399	16	Q7TVM3
37	126	6.2	426	16	Q7V508
38	124	6.1	411	2	Q83WZ9
39	123.5	6.1	467	16	Q87ZG1
40	123.5	6.1	498	16	Q8PZ00
41	122.5	6.1	520	16	Q988D6
42	122	6.0	403	16	Q7V123
43	121	6.0	549	16	Q8PGP1
44	120	5.9	525	10	Q8L8H4
45	119.5	5.9	397	16	Q82PA6

ALIGNMENTS

RESULT 1

Q8GCS2 PRELIMINARY; PRT; 382 AA.
 ID Q8GCS2
 AC Q8GCS2;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Lycopene beta-cyclase.
 GN CRTV.
 OS Pantoea stewartii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=66269;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 8200;
 RA deSouza M.L., Kollmann S.R., Schroeder W.A.;
 RT "Carotenoid Biosynthesis (WO 02/079395 A2).";
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY166713; AAN85598.1;
 DR GO; GO:0045436; P:lycopene beta cyclase activity; IEA.
 DR GO; GO:0016117; P:carotenoid biosynthesis; IEA.
 DR InterPro; IPR008461; Crtv.
 DR Pfam; PF05897; Crtv; 1.
 SQ SEQUENCE 382 AA; 43257 MW; 7E93D4976CSFCCB1 CRC64;

Query Match 100.0%; Score 2021; DB 2; Length 382;
 Best Local Similarity 100.0%; Pred. No. 3.8e-160;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MOPHYDLILVAGLGLIALRLOOQHDPMDRILLITRAGPEAGNGHTWSFHEEDLTLMQHR	60
Db	1	MOPHYDLILVAGLGLIALRLOOQHDPMDRILLITRAGPEAGNGHTWSFHEEDLTLMQHR	60
QY	61	WIAPLVHEHNEFYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLMTHTAVSAVHAE	120
Db	61	WIAPLVHEHNEFYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLMTHTAVSAVHAE	120
QY	121	SVQLADGRITTHASTVIDGRGVTTPDSALRVGFGQAFIQEWOLSAFHLGLSFIIMDATVDQO	180

```
121 SVQLADGRITTHASTVIDGRGYTPDSALRVGFOAFICQEWQLSAPHGSLSPFIIMDATVDDQ 180
181 NGYRFVYTLPLSATALLIEDTHYIDKANLOAERARONIRDYAARQGWPLQTLREEQGAL 240
181 NGYRFVYTLPLSATALLIEDTHYIDKANLOAERARONIRDYAARQGWPLQTLREEQGAL 240
241 PITLTGDNROFWOQOQFACSGGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHOTI 300
241 PITLTGDNROFWOQOQFACSGGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHOTI 300
301 AHPAQORWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
301 AHPAQORWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
361 ILGKPPVPVFAALQAIMTTHR 382
361 ILGKPPVPVFAALQAIMTTHR 382

RESULT 2
47844
C Q47844 PRELIMINARY; PRT; 382 AA.
T 01-NOV-1996 (T-EMBLrel. 01, Created)
T 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Dycopene cyclase.
N CRTY.
S Pantoea agglomerans.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=549;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=94236237; PubMed=8180698;
A To K.Y., Lai S.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
A Chang Y.S., Liu S.F.;
T "Analysis of the gene cluster encoding carotenoid biosynthesis in
T Erwinia herbicola Ehob3.";
L Microbiology 140:331-339 (1994).
R EMBL: M90698; AAA21262.1; -.
R PR; S52585;
R GO: 0045436; P:lycopene beta cyclase activity; IEA.
R GO: 0016117; P:carotenoid biosynthesis; IEA.
R InterPro: IPR008461; Crty.
R Pfam: PF05897; Crty; 1.
Q SEQUENCE 382 AA; 43248 MW; A3A3197C91BB1D64 CRC64;

Query Match 84.9%; Score 1715; DB 2; Length 382;
Best Local Similarity 83.8%; Pred. No. 1.2e-134;
Matches 320; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

Y 1 MOPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWSPFHEEDLTINQHR 60
b 1 MRPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWSPFHEEDLTINQHR 60
Y 61 WAPLVVHPDYQVRFQRRHNVNGYCVTSRHFAGILRQOFGHLLWLTAYSAVHAESV 120
b 61 WAPLVVHPDYQVRFQRRHNVNGYCVTSRHFAGILRQOFGHLLWLTAYSAVHAESV 120
Y 121 SVQLADGRITTHASTVIDGRGYTPDSALRVGFOAFICQEWQLSAPHGSLSPFIIMDATVDDQ 180
b 121 AVRLNNGQVISASAVIDGRGYTPNSALNVGFOAFICQEWQLSAPHGSLSPFIIMDATVDDQ 180
Y 181 NGYRFVYTLPLSATALLIEDTHYIDKANLOAERARONIRDYAARQGWPLQTLREEQGAL 240
b 181 NGYRFVYTLPLSATALLIEDTHYIDKANLOAERARONIRDYAARQGWPLQTLREEQGAL 240
Y 241 PITLTGDNROFWOQOQFACSGGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHOTI 300
b 241 PITLTGDNROFWOQOQFACSGGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHOTI 300
```

```
301 AHPAQORWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
301 THEFAHERWQOQGFPRMLNRMFLAGPADSRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
361 ILGKPPVPVFAALQAIMTTHR 382
361 ILGKPPVPVFAALQAIMTTHR 382

RESULT 3
Q8VUJ6
ID Q8VUJ6 PRELIMINARY; PRT; 386 AA.
AC Q8VUJ6;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DB Crty protein.
GN CRTY.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OK NCBI_TaxID=182454;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamiunten H., Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801.";
RI Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79602.1; -.
DR GO: 0045436; P:lycopene beta cyclase activity; IEA.
DR GO: 0016117; P:carotenoid biosynthesis; IEA.
DR InterPro: IPR008461; Crty.
DR Pfam: PF05897; Crty; 1.
SQ SEQUENCE 386 AA; 43674 MW; 1E7DB42C08447AE7 CRC64;

Query Match 73.9%; Score 1494; DB 2; Length 386;
Best Local Similarity 73.4%; Pred. No. 3.4e-116;
Matches 279; Conservative 36; Mismatches 65; Indels 0; Gaps 0;

Y 3 PHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWSPFHEEDLTINQHRWI 62
b 2 PRYDLILVAGLANGLIALRLQOQPSLAILLIDAREEPGANHTWSPFHEEDLTINQHRWI 61
Y 63 APLVHHWHPDYQVRFQRRHNVNGYCVTSRHFAGILRQOFGHLLWLTAYSAVHAESV 122
b 62 APLVHHWHPDYQVRFQRRHNVNGYCVTSRHFAGILRQOFGHLLWLTAYSAVHAESV 121
Y 123 QLADGRITTHASTVIDGRGYTPDSALRVGFOAFICQEWQLSAPHGSLSPFIIMDATVDDQNG 182
b 122 TLDDGRVLESDAVIDGRGYQPDGALRMGFSFVGQEWQLSEPHGLTAPIMDATVDDQAG 181
Y 183 YRFVYTLPLSATALLIEDTHYIDKANLOAERARONIRDYAARQGWPLQTLREEQGALPI 242
b 182 YRFVYSLPFSADTLIEDTHYIDNATLEGDRARONIRDYAARQGWPLQTLREEQGALPI 241
Y 243 TLTDGDNROFWOQOQFACSGGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHOTIAH 302
b 242 TLTDGVAARFWORDLPCSGRLAGLFHPTTGYSLPLAVALADRLAQMOTFSTLHATIQO 301
Y 303 FAQRWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLRIL 362
b 302 FASQWQOQGFPRMLNRMFLAGPADQRQWQVQRFYGLPEGLIARFYAGKLTVDRLRIL 361
Y 363 SGKPPVPVFAALQAIMTTHR 382
b 362 SGKPPVPVFAALQAIMTPHR 381

RESULT 4
Q93C18
ID Q93C18 PRELIMINARY; PRT; 385 AA.
AC Q93C18;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
```

```

1 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
2 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
3 Lycopene cyclase.
4 CRTX.
5
6 Xanthobacter sp. (strain Py2).
7 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
8 Hypnobiaceae; Xanthobacter.
9 NCBI_TaxID=78245;
10 [1]_
11 SEQUENCE FROM N.A.
12 STRAIN=Py2;
13 Larsen R.A., Metcalf W.W.;
14 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
15 EMBL; AF408848; AAL01999.1; -.
16 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.
17 GO; GO:0016117; P:carotenoid biosynthesis; IEA.
18 InterPro; IPR008461; CrtY.
19 Pfam; PF05897; CrtY; 1.
20 SEQUENCE 385 AA; 41442 MW; 8F69C80ECF6509FE CRC64;
21
22 Query Match 44.5%; Score 899; DB 2; Length 385;
23 Best Local Similarity 48.3%; Pred. No. 1.4e-66;
24 Matches 182; Conservative 57; Mismatches 136; Indels 2; Gaps 1;
25
26 6 DLIVGAGLANGLIALRLQOQHPDMRILLIIBAGPEAGNHTWSPHEEDLTINQHRWIAPL 65
27 2 DIVFVAGLANCLMAARLAQRPGELMLLEAGSVGGNHSWCSHSDLTAAQRAFLAPF 61
28
29 66 VVHHPDQVPRRRHNSGYCVTSRHFAGILRQOFGQHLMTAVSAVHAESVOLA 125
30 62 QSLWAGHGVHPAPSRTIKGGYATISSERVAEVMNERLCARITNARVAHPDHVLE 121
31
32 126 DGRITIASTVIDRGYTPDSALRVGFQAFIQEWQLSAPHGLSSPIIMDATVDOQNGYRF 185
33 122 GGERIDASAADVGRPLASRHLDLGYOTFLGQELMSRPHGLTRPIIMDARVEQLGGRF 181
34
35 186 VYTLPLSATALLIEDTHYIDKAMQAEARQNIDYAAQGWMPLOTLLRBEQALPITL 245
36 182 VYVLEDDTLLVEDTYTADGPDLPFADALRGRI SAYAAQGWADYVVRSEGLPLALG 241
37
38 246 GDNROFWOQOPO--ACSGLRAGLFHPTTGYSLPLAVALADRLSADVFTSSSVHQTIAHF 303
39 242 GDINAFLETSGVAPAGLAGLFHPTTGYSLPDAMALADSVSALADLSGPAVAVRSH 301
40
41 304 AQORWQOQGFPMNLNMLFLAGPAESRWVQRYGCLPEDIARFYAGKLTVDRLRILS 363
42 302 AAAANGRGFFLLNRLFRADPERRYAILQRFYGLSEDIARFYADRLTLADKARILS 361
43
44 364 GKPPVPVFAALQAIMT 380
45 362 GRPPVSVFALSCLVET 378
46
47 5
48 KIX3
49 Q9KIX3 PRELIMINARY; PRT; 394 AA.
50 Q9KIX3
51 01-OCT-2000 (TrEMBLrel. 15, Created)
52 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
53 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
54 Lycopene cyclase.
55 CRTX.
56
57 Bradyrhizobium sp. ORS278.
58 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
59 Bradyrhizobiaceae; Bradyrhizobium.
60 NCBI_TaxID=114615;
61 [1]
62 SEQUENCE FROM N.A.
63 STRAIN=ORS278;
64 MEDLINE=20309720; PubMed=10851005;
65 Hannibal L., Loxquin J., Angles d'Ortoli N., Garcia N.,
66 Chaintreuil C., Masson-Bovivin C., Dreyfus B., Giraud E.;
67 "Isolation and characterization of the canthaxanthin biosynthesis
68 genes from the photosynthetic bacterium Bradyrhizobium sp. strain
69 ORS278".
70 J. Bacteriol. 182:3850-3853(2000).
71
72 EMBL; AF218415; AAF78200.1; -.
73 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.
74 GO; GO:0016491; F:oxidoreductase activity; IEA.
75 GO; GO:0016117; P:carotenoid biosynthesis; IEA.
76 GO; GO:0006118; P:electron transport; IEA.
77 InterPro; IPR008461; CrtY.
78 InterPro; IPR000172; GMC_oxred.
79 InterPro; IPR000437; Prok_lipoprot_S.
80 Pfam; PF05897; CrtY; 1.
81 Pfam; PF00732; GMC_oxred N; 1.
82 PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
83 SEQUENCE 394 AA; 43435 MW; 8BDD59304EB194A CRC64;
84
85 Query Match 39.8%; Score 804; DB 2; Length 394;
86 Best Local Similarity 43.9%; Pred. No. 1.2e-58;
87 Matches 168; Conservative 63; Mismatches 138; Indels 14; Gaps 6;
88
89 6 DLIVGAGLANGLIALRLQOQHPDMRILLIIBAGPEAGNHTWSPHEEDLTINQHRWIAPL 65
90 6 DIVVIGGLAGGLIALRLTDAEPOLRVVIEGSASIAAGNHTWSPFGTIDSSDQHAWLGR 65
91
92 66 VVHHPDQVPRRRHNSGYCVTSRHFAGILRQOFGQHLMLH-TAVSAVHAESVOL 124
93 66 VGRWPGVEVFAEHAIRLSTAYLSMTSTRLEAEVQEPFERILRDATATISAT-ADHYVL 124
94
95 125 ADGRITIASTVIDRGYTPDSALRVGFQAFIQEWQLSAPHGLSSPIIMDATVDOQNGYR 184
96 125 EGRTILAPCVUIDARGRPVGLALGFKFLGVRLAAPHGLDVPVINDATVAGSDGYR 184
97
98 185 FYTTLPLSATALLIEDTHYIDKAMQAEARQNIDYAAQGWMPLOTLLRBEQALPITL 244
99 185 FYTTLPLDPQRLIEDTYSDGGELPEQVLRQIARVALAKGQIAETIRAEQGVLPVIL 244
100
101 245 TCDNRQFQQO---QPQACSLRAGLFHPTTGYSLPLAVALAD---RLSALDVTSSSVH 297
102 245 AGDPGLSKVSDSPRV--GLAALLVHPTTGYSLPDVAVRADLLTARLAQAGLSSADAR 302
103
104 298 QTIHFAQORWQOQGFPMNLNMLFLAGPAESRWVQRYGCLPEDIARFYAGKLTVD 357
105 303 ETIDVGRTIERRGYGFRFLNRLFKAAEPSEERILARFYGLDQALIERFYAARIQOD 362
106
107 358 RLRI---LSGKPPVPVFAALQAI 377
108 363 KLRVFMHMLMKPPPISSALACL 385
109
110 RESULT 6
111 P94791 PRELIMINARY; PRT; 382 AA.
112 ID P94791
113 AC P94791;
114 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
115 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
116 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
117 DE Lycopene cyclase.
118 GN CRTX.
119 OS Bradyrhizobium sp. ATCC 21588.
120 OC Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales;
121 OC Flavobacteriaceae; Flavobacterium.
122 OK NCBI_TaxID=50286;
123 RN [1]
124 RP SEQUENCE FROM N.A.
125 RC STRAIN=R1534;
126 RX MEDLINE=97186694; PubMed=9034310;
127 RA Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
128 van Loon A.P.;
129 RT "Isolation and characterization of the carotenoid biosynthesis genes
130 of Flavobacterium sp. strain R1534.";
131 RL Gene 185:35-41(1997).
132 DR EMBL; U62808; AAC44851.1; -.
133 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.

```

R GO: 0016117; P: carotenoid biosynthesis; IEA.
 R InterPro: IPR008461; CrtY.
 R Pfam: PF05897; CrtY; 1.
 Q SEQUENCE 382 AA; 42369 MW; 6DB5452F45D9EF4B CRC64;
 Query Match 38.8%; Score 783.5; DB 2; Length 382;
 Best Local Similarity 43.7%; Pred. NO. 5.7e-57;
 Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
 Y 5 YDILVAGLANGLIALRLQQCHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHRWIA- 63
 b 3 HDLLIAGAGLSCALIALAVDRRDPARIWLDARGSPSDQHTWSCHDTLS---PEMLAR 59
 Y 64 --PLVVHHPDYQVRPQRSHVNSGYCVTSRHFAGILRQFGQHLWLTAVSAVHAES 121
 b 60 LSPIRGEHTDQEVAPDHSRLLTGYGIEAGAILGLQ---GVDELWNTHTVAFLDTG 116
 Y 122 VQLADGRIIHAFTVIDGRGVTDPDSALRVGFQAFIGQEWQLSAPHLSSPIIMDATVDQON 181
 b 117 ATLTDGSRIEACVIDARGAVETPLITVGQKFGVGEIETDAPHGVVERPMINDATVPQMD 176
 Y 182 GYRFVYTLPLSATALLIBETHIDKANLQARQNIQDYAARQGWPLQTLRLREOQALP 241
 b 177 GYRFYLLSPSPRIILIEDTRYSDGDLDDGALAQAASLDYAARRGWTQOB--MRRERGILP 235
 Y 242 ITLTGDNROFWOQPOAC--SGLRAGLFHTTGYSLPLAVALADRLSALDVFETSSVHQT 299
 b 236 IALAHDAIGFWRDHAGQAVPVGLGAGLFHTTGYSLPYAAQVADAIARDL--TTASABRA 294
 Y 300 IAHFAQRQWQOQGFEMLRNMLFLAGPABSRVWQRFYGLPEDIILAFYAGKLTVTDR 359
 b 295 VRGWAIDRADRDRLFLNMLFRGCPDPRYLLIQRFYRLPQPLIFRYAGRLTLADRL 354
 Y 360 RLISGKPPVPVFAALQAI 377
 b 355 RIVTGPPPLSQAVRCL 372
 RESULT 7
 9RLH5 .PRELIMINARY; PRT; 386 AA.
 C Q9RLH5
 T 01-MAY-2000 (T-EMBLrel. 13, Created)
 F 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 E Lycopene beta-cyclase.
 N CrtY.
 S Paracoccus marcusii.
 C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodocyclales;
 C Rhodocyclaceae; Paracoccus.
 X NCBI_TaxID=59779;
 N [1]
 P SEQUENCE FROM N.A.
 C STRAIN=MH1;
 A Harker M., Hirschberg J.;
 T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
 ML";
 L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 R EMBL; Y15112; CAB56061.1; --
 R GO: 0045436; F: lycopene beta cyclase activity; IEA.
 R GO: 0016117; P: carotenoid biosynthesis; IEA.
 R InterPro: IPR008461; CrtY.
 R Pfam: PF05897; CrtY; 1.
 Q SEQUENCE 386 AA; 42283 MW; 91952F0F33F36800 CRC64;
 Query Match 36.4%; Score 735.5; DB 2; Length 386;
 Best Local Similarity 41.5%; Pred. NO. 5.9e-53;
 Matches 158; Conservative 69; Mismatches 139; Indels 15; Gaps 8;
 Y 5 YDILVAGLANGLIALRLQQCHPDMRILLIE--AGPEAGNHTWSFHEEDLTINQHRWI 62
 b 3 HDVLLAGLANGLIALAARAARPDRLVLLDRAAGPSEG--HTWSCHDPDLS----PWL 57

QY 63 A---PLVVHHPDYQVRPQRSHVNSGYCVTSRHFAGILRQFGQHLWLTAVSAVHA 119
 DB 58 ARKLPRANWPDQEVFPFHARRLATGSGDGAALADAVVRSAGETRW--DSDIALIDA 116
 QY 120 ESQVADGRIIHAFTVIDGRGVTDPDSALRVGFQAFIGQEWQLSAPHLSSPIIMDATVDQ 179
 DB 117 QGATLSCGTREIAGSVLDGRGPHSRHLTLGPHKFLGVEIETDRPHGVPRPVINDGTVQ 176
 QY 180 QNGYRVYTLPLSATALLIEDTHYIDKANLQARQNIQDYAARQGWPLQTLRLREOQA 239
 DB 177 RDGYGIYLLPSPRIILIEDTRYSDGDLDDALAAASQDFARQGW--TGAEVREIRGI 235
 QY 240 LPTLTGDNROFWOQPOQ--ACSGLRAGLFHTTGYSLPLAVALADRLSALD--VFTSSSV 296
 DB 236 LPTALAHDAAGFWADHAEQFVPVGLRAGFFHTVGYSLPYAAQVADVVAGLSPGPTDAL 295
 QY 297 HQTIAHFAQRQWQOQGFEMLRNMLFLAGPABSRVWQRFYGLPEDIILAFYAGKLTVT 356
 DB 296 RGAIRDYAIADRARDRLFLNMLFRGCPDPRYLLIQRLYRMPHGLIERFYAGRLSVA 355
 QY 357 DRLRLISGKPPVPVFAALQAI 377
 DB 356 DQLRIVTGKPPPLGTAIRCL 376
 RESULT 8
 006756 .PRELIMINARY; PRT; 434 AA.
 ID C06756
 AC C06756
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DB Lycopene cyclase.
 GN CrtY.
 OS Erythrobacter longus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
 OC Sphingomonadaceae; Erythrobacter.
 OK NCBI_TaxID=1044;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=och 101;
 RX MEDLINE=97311406; PubMed=9168123;
 RA Matsumura H., Takeyama H., Husakabe E., Burgess J.G., Matsumaga T.;
 RT "cloning, sequencing and expression the carotenoid biosynthesis gene,
 RT lycopene cyclase and phycoene desaturase, from the aerobic
 RT photosynthetic bacterium Erythrobacter longus sp. strain och 101 in
 RT Escherichia coli";
 RL Gene 189;169-174 (1997).
 RL EMBL; D83513; BAA20275.1; --
 DR GO: 0045436; F: lycopene beta cyclase activity; IEA.
 DR GO: 0016117; P: carotenoid biosynthesis; IEA.
 DR InterPro: IPR008461; CrtY.
 DR Pfam: PF05897; CrtY; 1.
 SQ SEQUENCE 434 AA; 48018 MW; 9E440E09BA5AEF47 CRC64;
 Query Match 31.5%; Score 637; DB 2; Length 434;
 Best Local Similarity 38.9%; Pred. No. 1.1e-44;
 Matches 154; Conservative 61; Mismatches 149; Indels 32; Gaps 10;
 QY 6 DLTIVAGLANGLIALRLQQCHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHRWIAPL 65
 DB 16 DCAIVGGGLAGGLIALALQARPEFIRVIEAGRTTIGNHRWSWFDSDLSAGRALADP 75
 QY 66 VVHEWF--DYQVRFPPQRSHVNSGYCVTSRHF--AGILRQFGQHLWLTAVSAVHAESVQ 123
 DB 76 RQTDWEGGVEYRPPKYRKLKTAYRSMASTDHFHGLRALPESSVILGRKAVGLDARGVD 135
 QY 124 LADGRI-----IHAGTVIDGRGVTDPDSALRVGFQAFIGQEWQLSAPHLSSPIIMDATVD 178
 DB 136 LAPSGYQPATRINARSVIDCRSFKPSAHLKGGVFLGRHMLQEPHGVENPVIMDATVD 195
 QY 179 Q-----QNG--YRFVYTLPLSATALLIEDTHYIDKANLQARQNIQDYAARQGWPLQTL 232

```

b 196 QLAZPHGNGSYRFFVYVPLGSHDVFIEDTYADDPDLDRNALSGRIDQYARANGWENGTP 255
Y 233 LREEQALPITLTGDNRFQWQOPOQ-----ACSGLAGLPHPTTGYSLPLAV-----ALADR 284
b 256 VHEAGVLPV-LTGDFGFSAYQDEVRIPIGVAIGARGGFTPHITSTYMCVAVENALAMAEQ 314
Y 285 LSAIDVFTSSVHQITIAHF---AQRWQOQGFPRMLNMLFLAGPAESRWRVQRFYGLP 341
b 315 -----PDLSGEQLAAFFDSRRARHSXTGYVLLARFLFPFAAKPRKVKVQRFYGLR 367
Y 342 EOLIAFPYAGKLTVDRLAILLSKGPVPVPAALQAI 377
b 368 EGLIFRYAARSNTFDKRVLMGCEPPVAIHSAILAM 403

RESULT 9
Q9M424 PRELIMINARY; PRT; 498 AA.
D Q9M424;
I 01-OCT-2000 (TREMREL. 15, Created)
I 01-OCT-2000 (TREMREL. 15, Last sequence update)
I 01-OCT-2003 (TREMREL. 25, Last annotation update)
E Neoxanthin synthase.
S NXS.
S Solanum tuberosum (Potato).
S Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
S Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
S lamiales; Solanales; Solanaceae; Solanum.
S NCBI_TaxID=4113;
S [1]
S SEQUENCE FROM N.A.
S STRAIN=cv. Desiree; TISSUE=Tuber;
S MEDLINE=20547529; PubMed=11094161;
S A Al-Babili S., Hugueney P., Schledz M., Walsch R., Protnmeyer H.,
S Laule O., Beyer P.;
S "Identification of a novel gene coding for neoxanthin synthase from
S Solanum tuberosum";
S FEBS Lett. 485:168-172(2000).
S EMBL; AJ271336; CAB92977.1; -.
S InterPro; IPR008671; Lycopen_cycl.
S R InterPro; IPR000205; NAD_BS.
S Pfam; PF05834; Lycopen_cycl; 1.
S SEQUENCE 498 AA; 56356 MW; 0CAB552D958C6F26 CRC64;

Query Match 7.7%; Score 156.5; DB 10; Length 498;
Best Local Similarity 22.1%; Pred. No. 0.00016;
Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

Y 5 YDLILVAGLAGLIALRLQOQ--QHPDMRILLIEAGPEA---GGNHTWSFHEEDLTIN-- 57
D 82 FDVILIGAGPA---GLRLAEHVSKYGIKVCVDPSPLSMFPNNYGVWVDFENLGEDC 137
Y 58 -QHRWIAPIVHHWPDYQVRF---PQRR-----RHVNSGYCVTSRHFAGILRQOFG 105
D 138 LDHKW---PMTCVHINDKTKYLRGPRYGRVSRKXKLRLNS---CVENR-----VKFYK 186
Y 106 QHLWLTAVSAVHAE---SVQLADGRIIHAFTVIDGRGYTP-----DSALRVGFQAFIGQ 157
D 187 AKVW-----KVEHEEFESIVCDDGKIRGSLVVDASGFASDFIEYDPRNHGYQIAHGV 241
Y 158 EQWL-SAPHGLSSPIIMD-----ATVQQNGYRFVYVTLPLSATALLIEDTHVID 205
D 242 LVEVDNHPDLDKWLMDWRDLSHLCNEPFLRVNNAKEPTFLYAMFDRNLVLETSLSV 301
Y 206 KANLAERARQNIIRYARQGWPLQTLREEGALPITLTGDNRFQWQOQFQACSGLRAG 265
D 302 RPLVSTMEVKRWRVARLHGLIKVKSVEIEKCVIPM---GGPLRIPQNVMAIGG-NSG 357
Y 266 LPHPTTGYSLPLAVALADRLS 286
D 358 IVHPSTGYVARSMAALAPVLA 378

RESULT 11
Q9XGX3 PRELIMINARY; PRT; 524 AA.
D Q9XGX3;
I 01-NOV-1999 (TREMREL. 12, Created)
I 01-OCT-2001 (TREMREL. 18, Last sequence update)
I 01-OCT-2003 (TREMREL. 25, Last annotation update)
E Lycopen cyclase.
S Citrus paradisi (Grapefruit).
S Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
S Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
S eurosids II; Sapindales; Rutaceae; Citrus.
S NCBI_TaxID=37656;
S [1]
S SEQUENCE FROM N.A.
S Costa M.C., Moreira C.D., Melton J.R., Otoni W.C., Moore G.A.;
S "Developmental expression of carotenoid genes in Citrus.";

```

```

RESULT 10
Q9FV32 PRELIMINARY; PRT; 498 AA.
D Q9FV32;
I 01-MAR-2001 (TREMREL. 16, Created)
I 01-MAR-2001 (TREMREL. 16, Last sequence update)
I 01-OCT-2003 (TREMREL. 25, Last annotation update)
E Chromoplast-specific lycopen beta-cyclase.
S Lycopersicon esculentum (Tomato).
S Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
S Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
S lamiales; Solanales; Solanaceae; Solanum.
S NCBI_TaxID=4081;
S [1]
S SEQUENCE FROM N.A.
S MEDLINE=20461502; PubMed=10995464;
S Ronen G., Carmel-Goren L., Zamir D., Hirschberg J.;
S "An alternative pathway to beta-carotene formation in plant
S chromoplasts discovered by map-based cloning of Beta and old-gold
S color mutations in tomato.";
S Proc. Natl. Acad. Sci. U.S.A. 97:11102-11107(2000).
S EMBL; AP254793; AAG21133.1; -.
S InterPro; IPR008671; Lycopen_cycl.
S DR InterPro; IPR000205; NAD_BS.
S Pfam; PF05834; Lycopen_cycl; 1.
S SEQUENCE 498 AA; 56448 MW; 2C21D3480746D5AA CRC64;

Query Match 7.5%; Score 152.5; DB 10; Length 498;
Best Local Similarity 22.1%; Pred. No. 0.00034;
Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

Y 5 YDLILVAGLAGLIALRLQOQ--QHPDMRILLIEAGPEA---GGNHTWSFHEEDLTIN-- 57
D 82 FDVILIGAGPA---GLRLAEHVSKYGIKVCVDPSPLSMFPNNYGVWVDFENLGEDC 137
Y 58 -QHRWIAPIVHHWPDYQVRF---PQRR-----RHVNSGYCVTSRHFAGILRQOFG 105
D 138 LDHKW---PMTCVHINDKTKYLRGPRYGRVSRKXKLRLNS---CVENR-----VKFYK 186
Y 106 QHLWLTAVSAVHAE---SVQLADGRIIHAFTVIDGRGYTP-----DSALRVGFQAFIGQ 157
D 187 AKVW-----KVEHEEFESIVCDDGKIRGSLVVDASGFASDFIEYDPRNHGYQIAHGV 241
Y 158 EQWL-SAPHGLSSPIIMD-----ATVQQNGYRFVYVTLPLSATALLIEDTHVID 205
D 242 LVEVDNHPDLDKWLMDWRDLSHLCNEPFLRVNNAKEPTFLYAMFDRNLVLETSLSV 301
Y 206 KANLAERARQNIIRYARQGWPLQTLREEGALPITLTGDNRFQWQOQFQACSGLRAG 265
D 302 RPLVSTMEVKRWRVARLHGLIKVKSVEIEKCVIPM---GGPLRIPQNVMAIGG-NSG 357
Y 266 LPHPTTGYSLPLAVALADRLS 286
D 358 IVHPSTGYVARSMAALAPVLA 378

RESULT 11
Q9XGX3 PRELIMINARY; PRT; 524 AA.
D Q9XGX3;
I 01-NOV-1999 (TREMREL. 12, Created)
I 01-OCT-2001 (TREMREL. 18, Last sequence update)
I 01-OCT-2003 (TREMREL. 25, Last annotation update)
E Lycopen cyclase.
S Citrus paradisi (Grapefruit).
S Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
S Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
S eurosids II; Sapindales; Rutaceae; Citrus.
S NCBI_TaxID=37656;
S [1]
S SEQUENCE FROM N.A.
S Costa M.C., Moreira C.D., Melton J.R., Otoni W.C., Moore G.A.;
S "Developmental expression of carotenoid genes in Citrus.";

```


Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

EMBL: AF152246; AAD38049.2; -
InterPro: IPR008671; Lycopene_cycl.
InterPro: IPR00205; NAD_BS.
Pfam: PF05834; Lycopene_cycl; 1.
SEQUENCE 524 AA; 58553 MW; 51823CF5B39208C CRC64;

Query Match 7.4%; Score 150; DB 10; Length 524;

Best Local Similarity 21.9%; Pred. No. 0.00058;

Matches 88; Conservative 66; Mismatches 162; Indels 86; Gaps 20;

6 DLILVAGLANGLIALRLQQQHPDMRILLIEAGPE---ACGNHTW---SFHEEDLTINQH 59

109 DLAVVGGGPGAG--LAVAQVSEAGLSVCSIDPSKLIWPNNGYVWVDFEAMDLDCDLD 166

60 RWIAPLVVHEWPDYQVRPPQRRRHVNSGYCVTSRHF--AGILRQQFQOHLWLHTA--VSA 116

167 TWSGAVV--HIDD-----NTKDLNRPYGRVNRKLLSKMLQKCTNGVKFQAKVIK 218

117 VHAESVQLA---DGRIIHASTVIDGRG-----YTPDSALRVGFOAFIGQEWQLS 162

219 IHEESKSLIICNDGVTIQAAVVLDTATGFSRCLVQYDKPNPGYQVAYGILAEVEQH--- 274

163 APGLSSPIIMD-----ATVDOQNGY--RFVYTLPLSATALLIEDTHYDKANQA 211

275 --PFOLDKRVFMDRSHLNNSQLAKANSKIPTFLYAMPFSSNRIFLZETSILVARPGVPM 333

212 BRASQNIIDYAAARQGWPLQTLREEQ-----GALPITLTGDNRFQWQQOPOACSGL--R 263

334 KDIOERWVARELKLGIKVKSEEDHCVPWGGPLVL-----PQRWVGIGT 381

264 AGLFHTPTGYSPLAVALADELS-----ALDVTSSSVHQTIAHFAQQRW-----QQQGF 313

382 AGWHPSTGYMVAVTLAARPIANAVRSLSORSISGRKLSAEVWKDLWPIERRQREF 441

314 FRM-INRMFLAGPAESRWVQRFYGLPEDLIARFYAGKLT 354

442 PCFGMDILLKDLPA-----TRFFDAFDELPYHGHFLS 477

RESULT 12

ID Q84VG9

AC Q84VG9

DT Q1-JUN-2003 (TrEMBLrel. 24, Created)

DT Q1-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT Q1-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Lycopene beta-cyclase.

EN PSI.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PACAD clade; Panicoideae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. W22;

RX MEDLINE=22558484; PubMed=12671084;

RA Singh M., Lewis P.E., Hardeman K., Bai L., Rose J.K., Mazourek M.,

Chonot P., Brutnell T.P.

RT "Activator Mutagenesis of the Pink scutellum1/viviparous7 Locus of

Maize."

RL Plant Cell 15:874-884 (2003).

DR EMBL: AY206862; AA018661.1; -

DR InterPro: IPR00205; NAD_BS.

SQ SEQUENCE 490 AA; 53360 MW; 05456FE5D355D1A4 CRC64;

Query Match 7.4%; Score 149; DB 10; Length 490;

Best Local Similarity 21.6%; Pred. No. 0.00064;

Matches 96; Conservative 69; Mismatches 164; Indels 116; Gaps 25;

6 DLILVAGLANGLIALRLQQQHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHRI--- 62

69 DLAVVGGGPGAGLAVARVAE--AGLSVCAIDPSPAV---VW-----PNNTYGVWVDEF 115

63 APLVVHH-----WPDYQVRPPQRRRHVNSGYCVTSRHFAGILRQQFQOHLW----- 109

116 EAMGLSRCLDTVWPSASV-----FIDGGAKSLORPYARVARRKLKSTMDRCVANGV 168

110 -LHTA--VSAVHAESVQLA---DGRIIHASTVIDGRGTP-----DSALRVGFOAFIGOE 158

169 VFHQAKVAKAVHYDASSLLICDDGVAVPASVLDATGFSRCLVQYDKPNPGYQVAYGIL 228

159 WQLSA--PHGLSSPIIMD-----ATVDOQNGY--RFVYTLPLSATALLIEDTHYDK 206

229 AEVDAHPFDIDKLMFMDRSHLPEGSEIRERNRRIPTFLYAMPFSPTRIFLETSILVAR 288

207 ANLOAESARQNIIDYAAARQGWPLQTLREEQ-----GALPITLTGDNRFQWQQOPOACS 260

289 PGLAMDDIQERMAARLRHLGIRVRSVEEDERCVPWGGPLVL-----PQRVV 336

261 GL--RAGLFHTPTGYSPLAVALADRL--SALDVTSSSV-----HQTIAHFAQQRW 308

337 GIGTAGKRVHPSTGYMVAVTLAARPIANAVRSLSORSISGRKLSAEVWKDLW 396

309 -----QQQGFPMRLNRMFLAGPAESRWVQRFYG-----LPEDL---I 345

397 PANRRQREF--CFGMDVLLKLDLGGTRFRFFDAFDELPYHGHFLSSRLFLPELLMFL 455

346 ARFYAGKLTVTDRILRLISGKEPPVPV 370

456 ALP--GNASNSRLEIMA--RGTVPL 477

RESULT 13

ID Q9M546

AC Q9M546

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Lycopene beta-cyclase.

OS Citrus sinensis (Sweet orange).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosid II; Sapindales; Rutaceae; Citrus.

OX NCBI_TaxID=2711;

RN [1]

RP SEQUENCE FROM N.A.

RA Xu C.J., Zhang S.L.;

RT "Molecular cloning of lycopene beta-cyclase gene from orange (Citrus

sinensis).";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF240787; AAP44700.2; -

DR InterPro: IPR008671; Lycopene_cycl.

DR InterPro: IPR00205; NAD_BS.

DR Pfam: PF05834; Lycopene_cycl; 1.

SQ SEQUENCE 504 AA; 56460 MW; 7C12B2BFB464481C CRC64;

Query Match 7.4%; Score 149; DB 10; Length 504;

Best Local Similarity 21.5%; Pred. No. 0.00067;

Matches 86; Conservative 64; Mismatches 168; Indels 82; Gaps 17;

6 DLILVAGLANGLIALRLQQQHPDMRILLIEAGPE---ACGNHTW---SFHEEDLTINQH 60

89 DLAVVGGGPGAG--LAVAQVSEAGLSVCSIDPSKLIWPNNGYVWVDFEAMDLDCDLD 146

61 WIAPLVVHEWPDYQVRPPQRRRHVNSGY-----CVTSRHFAGILRQQFQOHLWLHTA 113

147 TWSGAVVHHDDTKDLDRPYGRVNRKLLSKMLQKCTN-----GVKPHQAKV 195

114 VSAVHAESVQLA---DGRIIHASTVIDGRGTP-----DSALRVGFOAFIGQEWQLSA-P 164

196 IKVTHEESKSLIICNDGVTIQAAVVLDTATGFSRCLVQYDKPNPGYQVAYGILAEVEHP 255

165 HGLSSPIIMD-----ATVDOQNGY--RFVYTLPLSATALLIEDTHYDKANQOER 213

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	1678	83.0	382	1	CRTY_PANAN	P21687 pantoea ana
	2	1148.5	56.8	386	1	CRTY_ERWHE	Q01331 erwinia her
	3	759.5	37.6	386	1	CRTY_AGRAU	P54974 agrobacteri
	4	156	7.7	503	1	CCS_GITSI	Q9sea0 citrus sine
	5	152	7.5	526	1	LCVE_LYCES	Q65937 lycopersico
	6	1483.5	7.3	411	1	LCVB_SYN7	Q55276 synecchococ
	7	148.5	7.3	498	1	CCS_CAPAN	Q42435 capsicum an
	8	148.5	7.3	501	1	LCVE_ARATH	Q38933 arabidopsis
	9	141.5	7.0	503	1	LCVB_NARPS	Q40424 narcissus p
	10	140	6.9	500	1	LCVB_TORAC	Q43578 nicotiana t
	11	139	6.9	498	1	LCVB_CAPAN	Q43415 capsicum an
	12	133	6.6	500	1	LCVB_LYCES	Q43503 lycopersico
	13	129.5	6.4	524	1	LCVE_ARATH	Q38932 arabidopsis
	14	113.5	5.6	428	1	FXIC_ECOLI	P31575 escherichia
	15	106.5	5.3	410	1	Y801_DETRA	Q9rwe8 deinococcus
	16	101.5	5.0	391	1	SPMD_ECOLI	P75728 escherichia
	17	98	4.8	867	1	GFMD_ECOLI	P77468 escherichia
	18	97	4.8	610	1	GLM3_THIPE	Q56275 t glucosami
	19	95.5	4.7	490	1	MET3_SCHPO	P78937 schizosacch
	20	95.5	4.7	829	1	PQOP_PSEFL	P55174 pseudomonas
	21	95	4.7	328	1	THI2_SCHPO	P40398 schizosacch
	22	94.5	4.7	1160	1	DPJ3_ECOLI	P10443 escherichia
	23	94	4.7	456	1	YRDJ_RHOSH	Q01856 rhodobacter
	24	94	4.7	496	1	MOO_PROMM	Q7v8e6 prochloroco
	25	94	4.7	502	1	MOO_SYNPK	Q7u517 synecchococ
	26	93.5	4.6	961	1	ATCU_VERPE	Q8zca7 versinia pe
	27	93	4.6	580	1	NABD_STROO	Q9x8n8 streptomyce
	28	92.5	4.6	632	1	EPFD_SCHPO	P87111 s probabie
	29	91	4.5	339	1	LGPI_HUMAN	P33441 homo sapien
	30	91	4.5	435	1	FXIC_BRAJA	P10331 bradyrhizob
	31	90.5	4.5	1460	1	DPJ3_SALTY	P14567 salmomella
	32	90	4.5	497	1	MOO_PROMP	Q7v2q2 prochloroco
	33	90	4.5	724	1	GLB1_XANCP	Q8pe48 xanthomonas

b 121 SVPLKGGVIGARVDRGGRYAANSALSVGFQAPFQGEWRLSHPHCLSSPIIMDATVDQ 180
 y 181 NGRFVYVPLSNTALLIEDTHYIDKANIQAERARONIRDYAARGWPLQTLRBEQAL 240
 b 181 NGRFVYVPLSNTALLIEDTHYIDKANIQAERARONIRDYAARGWPLQTLRBEQAL 240
 y 241 PTLTGDNQRFQWQOQPCAGLRFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 300
 b 241 PTLTGDNQRFQWQOQPCAGLRFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 300
 y 301 AHAQORWQOQGFRLMRLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRLR 360
 b 301 THEARWQOQGFRLMRLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRLR 360
 y 361 ILSGKPPVPVFAALQAIMTHR 382
 b 361 ILSGKPPVPVFAALQAIMTHR 382

RESULT 2
 RTY ERWHE STANDARD; PRT; 386 AA.
 C Q01331; 1993 (Rel. 25, Created)
 T 01-APR-1993 (Rel. 25, Last sequence update)
 T 10-OCT-2003 (Rel. 42, Last annotation update)
 E Lycopene cyclase.
 N Crty.
 S Erwinia herbicola.
 C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 C Enterobacteriaceae; Pantoea.
 X NCBI_TaxID=549;
 [1]
 P SEQUENCE FROM N.A.
 C STRAIN=ERO10;
 X MEDLINE=931138098; PubMed=8422926;
 A Hundle B.S., O'Brien D.A., Beyer P., Kleinig H., Hearst J.E.;
 T "In vitro expression and activity of lycopene cyclase and
 T beta-carotene hydroxylase from Erwinia herbicola.";
 L FEBS Lett. 315:329-334(1993).
 C -!- FUNCTION: Catalyzes the cyclization reaction which converts
 C lycopene to beta-carotene.
 C -!- PATHWAY: Carotenoid biosynthesis.
 C -----
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C -----
 C EMBL; M87280; AAA64980.1; -;
 C PIR; S52981; S52981.
 C InterPro; IPR008461; Crty.
 C Pfam; PF05897; Crty; 1.
 C Carotenoid biosynthesis.
 C SEQUENCE 386 AA; 43341 MW; 54A40563BFCFA980 CRC64;

Query Match 56.8%; Score 1148.5; DB 1; Length 386;
 Best Local Similarity 58.3%; Pred. No. 7.4e-88;
 Matches 218; Conservative 53; Mismatches 102; Indels 1; Gaps 1;
 y 6 DLIIVGAGLGLIALRLQOQHDMRILLIEAGPEAGNHTWSPFHEEDLTLMQHRWIAPL 65
 b 3 DLIIVGAGLGLIALRLQOQHDMRILLIEAGPEAGNHTWSPFHEEDLTLMQHRWIAPL 62
 y 66 VVHEWPDYQVRFPORRHVNSGYCYVTSRHFAGILQCFQGHLLHTAVSAVHA 125
 b 63 VAHAWPGVEYQFDPRLRLRARGVYSISRFAALHQAALGENIWLNCSEVPLPNSVRLA 122
 y 126 DGRIRHASTVIDGRGYTPDSALRVGFQAFIQGBWQLSAPHGLSSPIIMDATVDQ 185

Db 123 NGEALLAGAVIDRGVGTASSAMQTYQLFLGQQWRLTQPHGLTVPILMDATVAOQGYRF 182
 Qy 186 VYTLPLSATALLIEDTHYIDKANIQAERARONIRDYAARGWPLQTLRBEQALPITLT 245
 Db 183 VYTLPLSATALLIEDTHYIDKANIQAERARONIRDYAARGWPLQTLRBEQALPITLT 242
 Qy 246 GDNQRFQWQOQPCAGLRFHPTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFA 304
 Db 243 GDIQALWADAPGVPVSGMRAGLRFHPTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFA 302
 Qy 305 QRFQWQOQGFRLMRLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRLRILSG 364
 Db 303 ERWRWQOQGFRLMRLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRLRILSG 362
 Qy 365 KPPVPVFAALQAIM 378
 Db 363 KPPVPVFAALQAIM 376

RESULT 3
 ID CRTY AGRU STANDARD; PRT; 386 AA.
 AC P54974;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lycopene cyclase.
 GN CRTY.
 OS Agrobacterium aurantiacum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OX NCBI_TaxID=44155;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96062243; PubMed=7592436;
 RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,
 RA Ohtani T., Miki W.;
 RT "Structure and functional analysis of a marine bacterial carotenoid
 RT biosynthesis gene cluster and astaxanthin biosynthetic pathway
 RT proposed at the gene level.";
 RL J. Bacteriol. 177:6575-6584(1995).
 C -!- FUNCTION: Catalyzes the cyclization reaction which converts
 C lycopene to beta-carotene.
 C -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
 C biosynthetic pathway.
 C -----
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C -----
 C EMBL; D58420; BAA09593.1; -;
 C InterPro; IPR008461; Crty.
 C Pfam; PF05897; Crty; 1.
 C Carotenoid biosynthesis.
 C SEQUENCE 386 AA; 42202 MW; B14C41B688AF78AC CRC64;

Query Match 37.6%; Score 759.5; DB 1; Length 386;
 Best Local Similarity 43.0%; Pred. No. 1.4e-55;
 Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;
 Qy 5 YDILVAGLGLIALRLQOQHDMRILLIEAGPEAGNHTWSPFHEEDLTLMQHRWI 62
 Db 3 HDVLVAGLGLIALRLQOQHDMRILLIEAGPEAGNHTWSPFHEEDLTLMQHRWI 57
 Qy 63 A---PLVHEWPDYQVRFPORRHVNSGYCYVTSRHFAGILQCFQGHLLHTAVSAVHA 119
 Db 58 ARUKPLERANWPDQVRFPORRHVNSGYCYVTSRHFAGILQCFQGHLLHTAVSAVHA 116
 Qy 120 ESQVLADGRIIHAHVTDGRGYTPDSALRVGFQAFIQGBWQLSAPHGLSSPIIMDATVDQ 179

117 QGATLSCGTEIAGAVLDGGAQPSRLTVGQKFGVGTETDRPAGVPRFMIMATVQ 176
 180 QNGYRFFVYPLSATALLIEDTHYIDKAMQAEARQNRDVAARQGMPLQTLLEBOQA 239
 177 QDGYRFFVYPLSATALLIEDTHYIDKAMQAEARQNRDVAARQGMPLQTLLEBOQA 239
 177 QDGYRFFVYPLSATALLIEDTHYIDKAMQAEARQNRDVAARQGMPLQTLLEBOQA 235
 240 LPTLTGDNQRQWQQP--QACSLGRLGPHPTTGYSLFLAVLADRLSALD-VFTSSSV 296
 236 LPTLAHDAAGFWADHAAGVPLGRLGPHPTTGYSLFLAVLADRLSALD-VFTSSSV 295
 297 HOTIAHFAQORWQQGFFRMLNMLFLAGPAESRWVQRVYCLPDLIARFYAGKLTVT 356
 296 RGAIRDYATDRARDRLNMLFRCGAPDRRYTLQRFYRPMHGLIERFYAGRLSVA 355
 357 DRLRLSGKPPVFAALQAI 377
 356 DQURIVTGKPPPLGTAIRCL 376

RESULT 4

CS_CITSI STANDARD; PRT; 503 AA.

D_QCS_CITSI STANDARD; PRT; 503 AA.
 Q953A0;
 16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Capsanthin/capsorubin synthase, chloroplast precursor.
 CCS.
 Citrus sinensis (Sweet orange).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 eurosids II; Sapindales; Rutaceae; Citrus.
 NCBI_TaxID=2711;
 [1]
 SEQUENCE FROM N.A.
 Xu C.J., Chen D.M., Zhang S.L.;
 "Molecular cloning of capsanthin/capsorubin synthase gene from orange
 (Citrus sinensis).";
 Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.
 -/- FUNCTION: CATALYZES THE CONVERSION OF THE UBIQUITOUS 5,6-
 EPOXYCAPSOTENIDS, ANTHEPAXANTHIN AND VIOLAXANTHIN, INTO CAPSANTHIN
 AND CAPSORUBIN, RESPECTIVELY.
 -/- PATHWAY: Carotenoid biosynthesis.
 -/- SUBCELLULAR LOCATION: Chloroplast; chromoplast (By similarity).
 -/- SIMILARITY: Belongs to the lycopene cyclase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF169241; AAF16189.1;
 InterPro; IPR008671; Lycopene_cycl.
 Pfam; PF05834; Lycopene_cycl.1.
 Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
 TRANSIT PEPTIDE.
 I CHAIN ? 503 CAPSANTHIN/CAPSORUBIN SYNTHASE.
 NP BIND 88 117 NAD (POTENTIAL).
 Q SEQUENCE 503 AA; 56663 MW; 2C957F0AA91075E6 CRC64;

Query Match 7.7%; Score 156; DB 1; Length 503;
 Best Local Similarity 21.2%; Pred. No. 2.2e-05;
 Matches 94; Conservative 69; Mismatches 166; Indels 114; Gaps 23;
 5 YDLJLVGAGLGLALRLQQQHPDWRLILLIAGPRA---GGNHTWSFHEEDLTINQHRW 61
 86 YDVIITGTFAGLRLAEQVSSRH-SVKVCVDPSFLSTWPNNGYVWVDFEDIGL----- 139

QY 62 IAPLVVHWP-----DYQVRFPPRRHVNQYCVTSRHF--AGILRQFOGHLWLT 112
 Db 140 -VCLDKTWPTCVFINDHKTKYDRP-----YGRVSRNLIKTKLENCVSNQGVFKH 191
 QY 113 AV--SAVHAB--SVQLADGRIIHAIVTIDGRGYTP-----DSALRVGFQAFITGOEWQL- 161
 Db 192 AKVMVNHQDFESSIVCDGNEIKASLIVDASGFASSFVEYDKPRNHGYQIAHGILAEVE 251
 QY 162 SAPHGLSPPIIMDATVDQONGYR-----FVYTLPLSATALLIEDTHYIDKANL 209
 Db 252 SHFPDLDKMWLDWR--DSHLGNBFIYRASNLKLPFLYAMPFDSNLVPLETSLVSRPVL 310
 QY 210 QASRARNIDYAAQWPLQTLLEBEQ-----GALPITLTGDNRFQWQQPQA--CSG 261
 Db 311 SYKEVKSRAARLRHNGIRVQRVIEDEKCLIPWGGLPVI-----PQSWAIG 358
 QY 262 LRAGLFHPTGY-----SLPLAVMALDR-----LSALDVFTSSSVHQTIAHFAQORWQQGF 313
 Db 359 GTSGLIHPATGYMVAFTMALAPALADAIACELGSTRMIRGRLHQKV-----W--NGL 409
 QY 314 F--RMLNMLFLAGPAESRWV---MORFYGLPEDLIARFVAGKLTVDRLRLISG-- 364
 Db 410 WPIDRRCNREFYSFG-METILLKLDLXGTRFPDAPFDLNPYYWHGFLSSRLSLAELAGLS 468
 QY 365 -----KPPVPV 370
 Db 469 LSLFGHASNSSRLDIIVTKCPVPL 491

RESULT 5

ID_LCYE_LCYE STANDARD; PRT; 526 AA.
 AC O65837;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lycopene epsilon cyclase, chloroplast precursor (EC 1.14.-.-).
 GN CRT1-E-1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. V736; TISSUE=Leaf;
 RA Ronen G., Cohen M., Zamir D., Hirschberg J.;
 RT "Regulation of expression of the gene for lycopene epsilon cyclase
 during fruit ripening of tomato."
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBSJ databases.
 CC -/- FUNCTION: CATALYZES THE SINGLE CYCLIZATION REACTION WHICH CONVERTS
 CC LYCOPENE TO DELTA-CAROTENE AND NEUROSPORENE TO ALPHA-ZEACAROTENE.
 CC -/- PATHWAY: Carotenoid biosynthesis.
 CC -/- SUBCELLULAR LOCATION: Chloroplast.
 CC -/- SIMILARITY: Belongs to the lycopene cyclase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; Y14387; CAA74745.1;
 DR PIR; T07082; T07082.
 DR InterPro; IPR001327; FAD pyr_redox.
 DR InterPro; IPR008671; Lycopene_cycl.
 DR InterPro; IPR001100; Pyr_redox.
 DR Pfam; PF05834; Lycopene_cycl.1.
 DR PRINTS; PR00368; FADPYR.
 DR PRINTS; PR00411; FNDPRTASE1.
 KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;

Transit peptide.
 1 ? CHLOROPLAST (POTENTIAL).
 2 ? 526 LYCOPENE EPSILON CYCLASE.
 3 ? 108 NAD (POTENTIAL).
 4 ? 136 NAD (POTENTIAL).
 5 ? 526 AA; 58886 MW; 37357C3869BDDCAF CRC64;
 6 DLIIIVAGLGLANGLIARLQQQHPDMRILLIEAGPE---AGNHTWSPHEBDLTNQHRI 62
 7 107 DLWVIGGPGAG---LALAAESAKULGNLGLV---GPDLPFTNNYGVWDEDFKDLGLQ----- 157
 8 63 APLVVHWPDPYQVRFPPRRRHVNSGYCYVTSRHF-----AGILRQOFGOHLWLH 111
 9 158 -ACIEHVWRDTIVYLDDEPILIGRAYGRVSRHFLHBEILLKRCVEAGV-----LYLN 208
 10 112 T-----AVSAVHAESVQLADRII---HASTVIDGRG-----YTPDSALRVGFOAFIGO 158
 11 209 SKVDRIYEATNGOSLVECEGDWIPCFVTVASGAASGKFLQYELGSP-RVSVQYAYGVE 267
 12 159 WQL-SAPHGSSPIIMDATVDQONGYR-----FVYTLPLSATALIIEDT 201
 13 268 VEVDNPPDPSLVFMD-----YEDYLRHDAQSLEAKYPTFLYAMPSPTRVFFET 319
 14 202 HYIDKANLQERARQNTDVAARQGWPLQTLLEBEOGALPITUTGNRFQWQOPOACSG 261
 15 320 CLASKDAMPFDLLKKMLRLNTLGVRAIKEIYEEBWSYIPVGSGLPN-----TEQKTLAFG 375
 16 262 LRAGLFHTTGYSLPLAVALADRLSAL-----DVFTSSSVHQTIAH-----FAQ 305
 17 376 AAASWVHATGYSVRSLSAPKASVLANILRQHYSKNMLTSSIPSISTQAMTWLWQ 435
 18 306 QRMQOQGFRRMLNMLFLAGPARRVWRVMOQFGLPDLIARFYAGKLTVD 357
 19 436 ERKQRSSF-LFGLIALQLDIEIGRSFFRAFRVPRKVMQGLGSLSSAD 486
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

DR EMBL; X74599; CAA52677.1; --
 DR InterPro; IPR008671; Lycopen cycl.
 DR InterPro; IPR000205; NAD BS.
 DR InterPro; IPR00103; Pyridine redox_2.
 DR Pfam; PF05834; Lycopen cycl; 1.
 DR PRINTS; PR00469; PNDRTASEII.
 KW Oxidoreductase; NAD; Carotenoid biosynthesis.
 FT NP_BIND 32 NAD (POTENTIAL).
 SQ SEQUENCE 411 AA; 46085 MW; C46CC5B2E85E7AC2 CRC64;
 Query Match 7.3%; Score 148.5; DB 1; Length 411;
 Best Local Similarity 22.7%; Pred No. 7.2e-05;
 Matches 97; Conservative 54; Mismatches 168; Indels 109; Gaps 20;
 QY 5 YDLIVAGLGLANGLIARLQQQHPDMRILLIEAGPEAGGNTWSPHEBDLTNQHRIAP 64
 DB 2 FDALVIGSGPAGLAAELAQ-----RGLKVQGLSPVDFPFWPE-----NTYGIWGP 48
 QY 65 -----LVVHHWPDYQVRFPPRRRHVNSGYCYVTSRHFAGIL-RQOFGOHLWLHTA- 113
 DB 49 ELDSLGLHFLGHRWNCVSYP-----GEAPVQHQNYGLFDRALQOH-WLRQCE 98
 QY 114 -----VSAV-----HAESVQLADRIIHAETVIDRGYTPDSALR-----VGFOAP 154
 DB 99 QGGLQWLQKAAIAHDSHSCVTTAAGQELQARLVDTTGHQAATFQRPHSDAIYQAA 158
 QY 155 IGEWQLS-----APHGSSPIIMDATVD-----QONGYRFVYTLPLSATALIETHYI 204
 DB 159 YGIIGQSQSPIEPHQF---VLMDFSDHLSPEERQUPPTFLYAMDGLNDVYFVETSLA 215
 QY 205 DKANLQERARQNTDVAARQGWPLQTLLEBEOGALPITUTGNRFQWQOPOACSLRA 264
 DB 216 ACPAIPYDLRLKQRLYQRLATRGVTVOVQIEEYCLFEMNLPLED-----LTQS VVGFGAA 271
 QY 265 GLFHTTGYSL-----PLAVALADRLSALDVFTSSSVHQTIAHFAQRNQ----- 309
 DB 272 SMVHPASGYMVGALLRAPDLANAIAGLNA-----SSSL--TTAELATQAMRGLWPTK 324
 QY 310 -----QCGFFRRMLNMLFLAGPARRVWRVMOQFGLPDLIARFYAGKLTVD-----R 358
 DB 325 IRKHYYVQGLEKL---MRESEALNHP---QTFGLPKQWYGLTNTLSLPELIQAM 378
 QY 359 LRILSGKP 366
 DB 379 LRIFAQAP 386
 RESULT 7
 CCS_CAPAN STANDARD; PRT; 498 AA.
 AC Q42435; Q39470;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Capsanthin/capsorubin synthase, chloroplast precursor.
 GN CCS.
 OS Capsicum annuum (Bell pepper).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Capsicum.
 CX NCBI_TaxID=4072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yolo Wonder;
 RX MEDLINE=94197697; PubMed=8147854;
 RA Duereux J., Bouvier F., Steppuhn J., Klein A., Camera B., Kuntz M.;
 RT "Structure and expression of two plant genes encoding chromoplast-
 specific proteins: occurrence of partially spliced transcripts.";
 RL Biochem. Biophys. Res. Commun. 199;1144-1150(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Lamuyo;
 RX MEDLINE=95004653; PubMed=7920703;


```

RL Nature 408:820-822(2000).
CC -!- FUNCTION: Catalyzes the double cyclization reaction which converts
CC lycopene to beta-carotene and neurosporene to beta-zeacarotene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the lycopene cyclase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U50739; AA053337.1; -.
DR EMBL; L40176; AAA81880.1; -.
DR EMBL; AF117256; AAF82388.1; -.
DR EMBL; AC009400; AAF02819.1; -.
DR InterPro; IPR008671; Lycopene_cycl.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF08833; Lycopene_cycl; 1.
DR OXIGREDUCTASE; NAD; Carotenoid biosynthesis; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 80 CHLOROPLAST (POTENTIAL).
FT CHAIN 81 501 LYCOPENE BETA CYCLASE.
FT NP_BIND 85 113 NAD (POTENTIAL).
FT CONFLICT 31 31 H -> P (IN REF. 2).
FT CONFLICT 243 243 V -> I (IN REF. 2).
FT SEQUENCE 501 AA; 56176 MW; C3014578D0BDC4E2 CRC64;
SQ
Query Match 7.3%; Score 148.5; DB 1; Length 501;
Best Local Similarity 21.08; Pred. No. 9.3e-05;
Matches 92; Conservative 66; Mismatches 175; Indels 105; Gaps 20;

QY 6 DLIIWAGIANGLIARLQOQHDPMRILLIAGPE---AGGNFTM--SFHEED-LTINQH 59
Db 84 DLIAIVGGGPAG--ILAVAQGVSEAGLSVCSIDPSFKLIWENNYGVWVDFEAMDLLDCLDT 141
QY 60 RRIALPLVHHWPDYQVRFQRRRHVNSGYCYVTSRHFAGILRQOF-----GQHL 108
Db 142 TWSGAVV-----FYDEGVKDLURPYGRVWRKQLKSKLOKCIITNGVKF 185
QY 109 MLHTAVSAVHAE---SVQLADGERIIHASTVIDRGYTP-----DSALRVGPAQAFIGQEMQ 160
Db 186 HQSKVTNVVHEEANSVTVCSGVKIQASVLDATGFSRCLVQYDKPYNPGVQVAYGVIAE 245
QY 161 LSA-PHGLSSPIINDATVDQNGY-----AFVYTLPLSATALLIEDTHYIDKAN 208
Db 246 VDGHEPFDYKVMFDMWRDKHLDYFELKERNSKIPTFLYAMPFSSNRIFLEETSLVARPG 305
QY 209 LQAEARQNIIDYAARQGWPLQTLRBEQ-----CALPITITGDNRQFWQQQQAQSG 262
Db 306 LEMEDIQERMAARLKHIGINVKRIEEDERCVIPWGGPLVL-----PQRVUGI 353
QY 263 --RAGLFHPTTGYSLPLAVALADRLS-----ALDVTSSSVH--QTIAHFAQQRW----- 308
Db 354 GGTAGWHPSTGYVMVARTLAAAPIVANAIVRYLGPSSNSLRGDSLAEVWRDLWPERR 413
QY 309 CQQGQFRMLNRLMFLAGPAESKRVRMORFGLPED-----LIARFVAGKLTV----- 355
Db 414 ROREFF-CFGMDILLKDLDATRAFFDAFDLQPHYWGCFLSRLPLFLVLLVFGLSLESH 472
QY 356 ---TDRLRILSKPPVPV 370
Db 473 ASNTSRLEIMT-KGTVPL 489

RESULT 9
LCYB_NARPS LCYB_NARPS STANDARD; PRT; 503 AA.
ID LCYB_NARPS
AC Q40424;
DT 16-OCT-2001 (Rel. 40, Created)

```

16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update) [BC 1.14, -.-].
Lycopene beta cyclase, chloroplast precursor [BC 1.14, -.-].
LCY1 OR LYC.
Narcissus pseudonarcissus (Daffodil).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
Narcissus.
NCBI_TaxID=39639;
[1]
SEQUENCE FROM N.A.
TISSUE=Paracolla;
Al-Babili S., Hobeika E., Beyer P.;
"A cDNA encoding lycopene cyclase from Narcissus pseudonarcissus L.";
(in) Plant gene Register PGR96-107.
[2]
SUBCELLULAR LOCATION.
MEDLINE=37433278; PubMed=9288918;
Bank M., Hoffmann B., von Lintig J., Schledz M., Al-Babili S.,
Hobeika E., Kleinig H., Beyer P.;
"Chloroplast import of four carotenoid biosynthetic enzymes in vitro
reveals differential fates prior to membrane binding and oligomeric
assembly.";
[3]
Eur. J. Biochem. 247:942-950(1997).
!- FUNCTION: Catalyzes the double cyclization reaction which converts
lycopene to beta-carotene and neurosporene to beta-zeacarotene.
!- PATHWAY: Carotenoid biosynthesis.
!- SUBCELLULAR LOCATION: Chloroplast; chromoplast. Exists as an
inactive soluble form and an active membrane-bound form
(probable).
!- SIMILARITY: Belongs to the lycopene cyclase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL: X98796; CAA67331.1; -.
InterPro: IPR009671; Lycopene cycl.
Pfam: PF05834; Lycopene_cycl_1.
Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
Transit peptide; Membrane.
TRANSIT 1 85 CHLOROPLAST (POTENTIAL).
CHAIN 86 503 LYCOPENE BETA CYCLASE.
NP_BIND 90 117 NAD (POTENTIAL).
SEQUENCE 503 AA; 56910 MW; 3FD1E355EF184D98 CRC64;

Query Match 7.0%; Score 141.5; DB 1; Length 503;
Best local Similarity 20.4%; Pred. No. 0.00036;
Matches 85; Conservative 66; Mismatches 150; Indels 115; Gaps 19;

QY 6 DLILVAG-----LANGLIALRLOQHPDMDRIILIEAGPEAGNHTWSPHFEDLTL 56
Db 89 DLAVVGGPLARSCSTSLGGLSVSD---PNPKL---WPNNYG--VWVDFEDMDL 139
QY 57 ---NQHRWIAFLVHHWPDYQVRFPPORRHVNSGYCVTSRHFAGILRQQF----- 104
Db 140 LDCLDATWSGAIV-----YVDRSTKSLSPYARVNRKLNKLSQWKKCV 183
QY 105 --GQHLWLHTAVSAVHAES---VQLADGRIHASVIDRGVTP-----DSALRVGQAF 154
Db 184 SNGVRFHQAATVKAHMEESKSYLICDGVTDIARVLDATGFSCLVQYKPYNGVQA 243
QY 155 IGQBWQLSA-PHGLSSPIIMD-----ATVQDQNGY--RFVYTLPLSATALLIEDTH 202
Db 244 YGILAEVEEHPFDVXWVFWMDRSHLNGKALNELNNAKIPTFLYAMPSSNRIFLETS 303
QY 203 YDKANTQAARQONTIDYARQGWPLQTLREO-----GALPITLTGDNRFWQQQP 256
Db 304 LVAREPGLKMDIQRVARLNHLGIRIKISIEDERCIVMPGGLPVI-----P 351

RESULT 9
LCYB_NARI
ID LCYB
AC Q40
DT 16-0


```

Y 257 QACSL--RAGLFHTTGYSLPLAVALADRLSALDVFSSVHQTIAH----- 302
b 352 QRAVIGGTAGMHPSTGY-----WVARTLAADIVANSIVQVLSVSDSLGENDLSADV 405
Y 303 ----FAQQRWQQGFRMLNRLFLAGPAESRWVQRFYGLPDLIARFYAGKLT 354
D 406 WIDLWPIERRRQREFF-CFEMDILLKDLLEGTRRFFDAFF-----DLSPRYWHGFLS 456

RESULT 10
CYB TOBAC
D LCYB TOBAC STANDARD; PRT; 500 AA.
Q43578;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
L Lycopene beta cyclase, chloroplast precursor [EC 1.14.-.-].
LCY1 OR CRT1.1.
Nicotiana tabacum (Common tobacco).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
lamids; Solanales; Solanaceae; Nicotiana.
NCBI_TaxID=4097;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Samsun NN; TISSUE=Leaf;
MEDLINE=96434545; PubMed=8837512;
Cunningham F.X. Jr., Pogson B., Sun Z., McDonald K.A., Dellapenna D.,
Gantt E.;
"Functional analysis of the beta and epsilon lycopene cyclase enzymes
of Arabidopsis reveals a mechanism for control of cyclic carotenoid
formation."
Plant Cell 8:1613-1626(1996).
-!- FUNCTION: Catalyzes the double cyclization reaction which converts
lycopene to beta-carotene and neurosporene to beta-zeacarotene.
-!- PATHWAY: Carotenoid biosynthesis.
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the lycopene cyclase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
EMBL; X81787; CAA57386.1; -.
PIR; S72506; S72506.
InterPro; IPR008671; Lycopene_cycl.
InterPro; IPR000205; NAD BS.
Pfam; PF05834; Lycopene_cycl; 1.
Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
Transit peptide.
TRANSIT 1 81 CHLOROPLAST (POTENTIAL).
CHAIN 82 500 LYCOPENE BETA CYCLASE.
NP_BIND 86 114 NAD (POTENTIAL).
SEQUENCE 500 AA; 56067 MW; 2E3721B87EE8CB8C CRC64;

Query Match
Best Local Similarity 6.9%; Score 140; DB 1; Length 500;
Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;

6 DLIVVAGLANGLIALRQQQHPDRILLIEAGPE---AGNHTW---SFHEED-LTLNQH 59
85 DLAVVGGGPAG--LAVAQVSEAGLSVVSIDPSKLIWPNNGVWVDFEAMDLDCLOA 142
60 RWIAPLVVHPDPVQRRPQRRHVNNSYCVTSRHPAGILRQQFGHLA-----L 110
143 TWSGTGVV-----VDNTTKDLDRPYGRVNRKQLSKMKQKILNGVKF 186
111 HTA--VSAVHAESVQLA---DGRITIHASTVIDGRGYTP-----DSALRVGFOAFIQGEWQ 160

```

```

187 HEAKVTKVTHEEAKSLCNDGVTTIQAATWLDATGFSRCLVQYDKPKGYQVAYGILAE 246
161 LSA-PHGLSSPIIMD-----ATVDOQNGY--RFVYTLPLSATALLIBDTHYIDKAN 208
247 VVEHPEDTTSKVLMDWRDHLGNMELKERNKVTFFLYAMPFSSNKIFLETSIVARPG 306
209 LQARARQNRIRYAARQGWPLQTLRREO-----GALPITLTGDNRFQWQOQOAC--S 260
307 LRMDDIQERWVARLNLGLIKVKSIEDEHCVIPMGSLFVI-----PQRVVGVT 354
261 GLRAGLFHTTGYSLPLAVALADRLSALDVFSSSVH-----QTLAHPAQQRW- 308
355 GGTAGLVHPSTGY-----WVARTLAADIVANSIVQVLSVSDSLGENDLSADV 408
309 ----QQQGFRLM-LNRMLFLAGPAESRWVQRFYGLPDLIARFYAGKLT 354
409 IERRRQREFFCFEMDILLKDLPA-----TERPFDADFLEPRYWHGFLS 453

RESULT 11
LCYB CAPAN STANDARD; PRT; 498 AA.
ID LCYB CAPAN
AC Q43415;
DF 16-OCT-2001 (Rel. 40, Created)
DF 16-OCT-2001 (Rel. 40, Last sequence update)
DF 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene beta cyclase, chloroplast precursor [EC 1.14.-.-].
GN LCY1 OR CRT1.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Lamuyo; TISSUE=Fruit;
MEDLINE=96045549; PubMed=7550379;
Hugueney P., Badillo A., Chen H.C., Klein A., Hirschberg J.,
Camara B., Kuntz M.;
"Metabolism of cyclic carotenoids: a model for the alteration of this
biosynthetic pathway in Capsicum annuum chromoplasts."
Plant J. 8:417-424(1995).
CC -!- FUNCTION: Catalyzes the double cyclization reaction which converts
lycopene to beta-carotene and neurosporene to beta-zeacarotene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- SIMILARITY: Belongs to the lycopene cyclase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X86221; CAA60119.1; -.
InterPro; IPR001327; FAD pyr redox.
InterPro; IPR008671; Lycopene_cycl.
InterPro; IPR000205; NAD BS.
InterPro; IPR001100; Pyr_redox.
Pfam; PF05834; Lycopene_cycl; 1.
PRINTS; PR00368; FADPNR.
PRINTS; PR00411; PNDRDTASEI.
KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW Transit peptide.
TRANSIT 1 79 CHLOROPLAST (POTENTIAL).
CHAIN 80 498 LYCOPENE BETA CYCLASE.
NP_BIND 84 112 NAD (POTENTIAL).
FT NP_BIND 84 112
SQ SEQUENCE 498 AA; 55610 MW; 177180CD5745F64F CRC64;

Query Match
Score 139; DB 1; Length 498;

```


"Gene structure and regulation of the carotenoid biosynthesis pathway in *Arabidopsis thaliana*.";
Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases

[31] SEQUENCE FROM N.A.
STRAIN=cv. Columbia;
MEDLINE=20181125; PubMed=10718197;
Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,
Tabata S.,
"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
features of the regions of 3,076,755 bp covered by sixty pl and TAC
clones.",
DNA Res. 7:31-53(2000).

[4] SEQUENCE FROM N. A.
STRAIN=cv, Columbia;
MEDLINE=22954850; PubMed=14593172;
Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
Chandra M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
Hayashizaki Y., Johnson-Kopson C., Hsuan V.W., Iida K., Karnes M.,
Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
Kamlya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
Yuan S., Shinoraki K., Davis R.W., Theologis A., Ecker J.R.:
"Empirical analysis of transcriptional activity in the Arabidopsis
genome".

Science 302:842-846(2003).

- I- FUNCTION: CATALYZES THE SINGLE CYCLIZATION REACTION WHICH CONVERTS LYCOPENE TO DELTA-CAROTENE AND NEUROSPORENE TO ALPHA-ZEACAROTENE.
- II- PATHWAY: Carotenoid biosynthesis.
- III- SUBCELLULAR LOCATION: Chloroplast.
- IV- SIMILARITY: Belongs to the lycopene cyclase family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/ebis/ebis-sib.ch>).

EMBL; U50738; AAB53336.1; --
EMBL; AF117257; AAF82389.1; --
EMBL; AB024035; BAA97033.1; --
EMBL; AY040024; BAA64181.1; --
EMBL; AY079371; AAL85102.1; --
InterPro; IPR001327; FAD pyr redox.
InterPro; IPR008671; Lycopene cycl.
Pfam; PF05834; Lycopene_cycl.
PRINTS; PR00368; FADENR.
Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
Transit peptide.
TRANSIT 1 45 CHLOROPLAST (POTENTIAL).
CHAIN 46 524 LYCOPENE ERSILON CYCLASE.
NP_BIND 111 139 NAD (POTENTIAL).
CONFLICT 111 111 L -> H (IN REF. 1).
SEQUENCE 524 AA; 58491 MW; 4C1F98CC72EDD074 CRC64;

```
Query Match          6.4%; Score 129.5; DB 1; Length 524;  
Best Local Similarity 20.0%; Pred. No. 0.0037;  
Matches      81; Conservative    65; Mismatches   176; Indels     83; Gaps  
  
        6 DULVAGAGLGLIALRLQQQHPDRIILLIISAPGE---ACGNHTWTSFHERDLTLN--QH  
           ||::|| ::| | :| ::| ::| ::| ::| ::  
       110 DLVVICGGPAG-LALAAESAKLGKLVGLI--GPDLPETNNYGVWEDEFNDLGLQCKIEH  
              :| ::| ::| ::| ::| ::| ::| ::| ::  
       60 RMIAPIVLVHHWPDYOV-----RFQRRRHWNSGYCVTSHFPAGILRQQFGHLWLHT  
            | ::| ::| ::| ::| ::| ::| ::| ::
```

```

166 VM-RETI VYLD DDKPITIGRAYGVSRRLHBE LLRCVES-----GV-----SYLSS 212
113 AVSAYHAE S-----VOLADGR IIEASTVIDRGYTPDSAL-----RVGFQAFIQEWQ 160
213 KVD S:TEASDGLRLVACDDNNVI PCRLATVASGAASKLLQYEVGGPRVCQTAYGVEVE 272
161 L-SAPHG LSSPIIMDA TVDQONGYR-----FVYTLPLSATALIEDTHY 203
273 VNSYSPYDDQWFM D-----YRDVNEKVSLEAEYPTFIYAMPMTKSRLPFBETCL 324
204 IDKANLQERARQNI RYAAQGMPLQTLREBQALPITLTDGNRQFMQQQPCASGLR 263
325 ASKQWMPD LTKLMRLD TLGIRILKTVEEBSWYIPVGGSLPN---TEQKNLAFGAA 380
264 AGLFHPPTGY SLPPLAVALADLSAL---DVFTSSSVHQTIAHFAQBSW-----QQQG 312
381 ASNVLPATGY SVTSLSEAPKYSVIAILREBETTKQINSIRQAWD TLWPPERKQRA 440
313 PFRMLNRML FLAGPABGRNWRVQSFYGLPBDL IARFYAGKLTATD 357
441 PF-LFGLALIVQFOTEGIRSFRTFFRLPKWMOGFLGSTLTSGD 484

```

RESULT 14
FIXC_ECOLI STANDARD; PRT; 428 AA.
ID FIXC_ECOLI
AC P31575; P75626;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE FIXC protein.
GN FIXC OR B0043 OR C0053.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
RN [1]_TaxID=562, 217992;
SEQUENCE FROM N.A.
RP
RC STRAIN=K12;
RX MEDLINE=923334977; PubMed=1630901;
RY Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RT Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RN Nucleic Acids Res. 20:3305-3308(1992).
RP
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RY Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RP
RC STRAIN=K12 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RY Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Zhou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Liou S.-R., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RN Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RP
RC STRAIN=O4:K74;
RX MEDLINE=96066354; PubMed=7473063;
RY Biehler K., Buchet A., Bourgis P., Kleber H.-P.,
RA Mandrand-Berthelot M.-A.;

"The fix Escherichia coli region contains four genes related to

carnitine metabolism".

J. Basic Microbiol 35:217-227(1995).

- FUNCTION: PROBABLY ACCEPTS ELECTRONS FROM FIXA/FIXE AND REDUCES A

QUINONE.

- COFACTOR: FAD (potential).

- SIMILARITY: Belongs to the BTP-QO / fixC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D10483; BAB96611.1; -
EMBL; AB000114; AAC73154.1; -
EMBL; AB016755; AAN78549.1; -
EMBL; X71977; CAA50799.1; -
PIR; C64725; C64725.
EcoGene; EGI1564; fixC.
InterPro; IPR001327; FAD pyr_redox.
InterPro; IPR000205; NAD_Bs
InterPro; IPR001100; Pyr_redox.
InterPro; IPR00103; Pyridine redox 2.
PRINTS; PR00368; FADPNR.
PRINTS; IPR003042; Rng_mnoxygenase.
PRINTS; PR00411; PNDROTASII.
PRINTS; PR00469; PNDROTASII.
PRINTS; PR00420; RNMNOXGNASE.
Oxidoreductase; FAD; Flavoprotein; Complete proteome.
NP BIND 12 28 FAD OR NAD(P) (ADP PART) (POTENTIAL).
CONFLICT 1 3 MSE -> MT (IN REF. 4).
CONFLICT 64 64 D -> E (IN REF. 1).
CONFLICT 280 281 VW -> IM (IN REF. 1).
CONFLICT 310 310 L -> F (IN REF. 1).
SEQUENCE 428 AA; 45702 MW; 4D089BESA30C7FBC CRC64;

Query Match 5.6%; Score 113.5; DB 1; Length 428;

Best Local Similarity 20.0%; Pred. No. 0.061;

Matches 88; Conservative 56; Mismatches 167; Indels 129; Gaps 17;

5 YDLIVGAGLGLIALRLQOHPDMRILLIEAGPAGGNNHTWSPHEBDLTLNQRWIAIP 64
6 FDLIVGAGLGLSVAALVLARE--GAQLVIERGNSAGAKN-----VTGSELVAH 53
65 LVVHWPDYQVRPPQRR--RHNVSQY-----YC-----VTSRHFAGILROOFGOH 107
54 SLEHIIICFADSAPVERLITHEKLAFTWSEKAMTNDYCNQDETSPQSRYSVLRSGF--D 111
108 LMLHTAVSAVHAESVQ-----LADRIIHASTVIDGRGTPDSALRVG 150
112 AMLMQEAEAGALITGIRVDNLVORDKGVGVGADGVIEAKTVILADGVNSILAEKIG 171
151 F-----QAFIQEWOLSAFGLSSPIINDATVDQONGYRFFVYILP 190
172 MAKRVKPTDVAVGKELIELEPKSVIEDRFQLOQNOG--AACLPAGSPDCLGWSGGFLYINE 230
191 LSATALLIEETHYIDKXNLQAEARQNRDYARQGWPL---QTLIREQGLP-----241
231 NTLISGLVLCGLHLDKAKSVPMQLEDFKQHPAVA--PLIAGKLVESAHVVPPEAGINM 288
242 -----ITLTGNRQFQWQQQACSGLEAGLFHTTGYSLPLVALADRLSALDVFETS 293
289 LPELVGQVLIAGD-----AAGCMNGLFTIRGMDLAIAGEAAKTVLSMKVS 337
294 SSVHQTAAHQQRWQQGFFRMNMLFLAGPESRWVRMQRVFLGPDLLTARFYAGKL 353
338 DD-----FSKQKLAEY-----RQHLESGLP-----EDMRMYQLPAFL-----370
354 TVTDRILSLGKPPVPVFAA 373

Db 371 --DNPRMFGYPPELAVGVA 387

RESULT 15

Y801 DEIRA

ID Y801 DEIRA STANDARD; PRT; 410 AA.

AC Q9RW68;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative carotenoid cyclase DR0801.

GN DR0801.

OS Deinococcus radiodurans.

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI_TaxID=1299;

RN SEQUENCE FROM N.A.

RP STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;

RC MEDLINE=20036896; PubMed=10567286;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

radiodurans R1.";

RL Science 286:1571-1577(1999).

CC - SIMILARITY: Belongs to the lycopene cyclase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB001934; AAF10377.1; -

PIR; D75475; D75475.

TIGR; DR0801; -

InterPro; IPR008671; Lycopene cycl.

InterPro; IPR003042; Rng_mnoxygenase.

Pfam; PF05834; Lycopene cycl. 1.

PRINTS; PR00420; RNMNOXGNASE.

KW Hypothetical protein; Oxidoreductase; NAD; Carotenoid biosynthesis;

Complete proteome.

NP BIND 11 39 NAD (POTENTIAL).

SEQUENCE 410 AA; 43161 MW; E1B3162F10F9E6AF CRC64;

Query Match 5.3%; Score 106.5; DB 1; Length 410;

Best Local Similarity 22.8%; Pred. No. 0.22; Indels 121; Gaps 22;

Matches 100; Conservative 44; Mismatches 173;

6 DLIVGAG-----LANGIALRLQOQ-----HPDMRILLIEAGPAGGNNHTWSPHEBDLTL 56

10 DVLIVGGSPGSLSAELAAAGLDVQLAPHP-RPPATYGAWLGLDPTWA-----60

57 NOHRWIAPIVHHWPDYQVRPPQRRHVSNGYCVTSRHFAGILROOFGHLLWHTAVSA 116

61 ---RGCAEQV---WTDVRAVYTPQPTSLGQVALLDN---AALLRTLRLGLADWTWVEGAA 111

117 VHAESVQLADGRIIHASTVIDGRGTPDSALRVGFQAFIQEWOL-----SAPHG-LSSP 170

112 LHAER-----SGAGWTYVGA-----GGERWQTVLVDASGHALVSP 148

171 IIMDATVDQONGY-----RFTYTLPLSATALL 197

149 VRFPFGAALQATAYGVVARFRPVPVTPGSMVMMDYRTPAPELKRGEATFLYAMHLGGDRYF 208

198 IEDTHYIDK-ANLQAEARQNRIDYARQGWPLQTLIREEQALPITLTGDNRFQWQOOP 256

209 VEETSLIARPAFTRAELERRLLARISA-QCTPPHATESEEWAFPM-----NAQAPAGG 262
257 QACSLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSSVHQTIAHFA--QORWOOQGF 314
263 VLAYGAAGRVRHVPVSGFQ--VAGALSDFGVATATATACQKDAAGWAALWSPE--- 317
315 RMLNRLFLAG-----PESRWVRVQRFYGLPEDLIARFY-----AGKLTVTDLRI 361
318 RRAAREVHLLGVGALLGLERAEPL-HFFGTFFGLPREQWARFLHPDPTDAGTLART-MLRV 375
362 LS---GKPPVVFPAALQA 376
376 FAQTGGVRVLPFLARALA 393

Search completed: February 29, 2004, 14:45:10
Job time : 10.4382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: February 29, 2004, 14:34:14 ; Search time 10.6328 Seconds
(without alignments)
3455.835 Million cell updates/sec

Title: US-09-941-947A-30

Perfect score: 2021

Sequence: 1 MQPHYDLILVAGLANGLIA.....SGKPPVPFAALQAIMTTHR 382

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB	ID	Description
1	1715	84.9	382	2	S52585	dycopen cyclase -
2	1678	83.0	382	2	C37802	crty protein - Erw
3	1148.5	56.8	386	2	S52981	lycopene cyclase -
4	152	7.5	526	2	T07082	lycopene epsilon-c
5	148.5	7.3	498	2	S71511	capsanthin/capsoru
6	147.5	7.3	471	2	S51511	lycopene beta-cycl
7	140	6.9	500	2	S72506	lycopene beta-cycl
8	139	6.9	500	2	S66349	lycopene beta-cycl
9	133	6.6	500	2	S72505	lycopene beta-cycl
10	133	6.6	500	2	S66350	lycopene beta-cycl
11	127.5	6.3	399	1	E70888	probable glf prote
12	117.5	5.8	391	2	AD0584	probable monooxyge
13	114.5	5.7	428	2	F0634	flavoprotein [impo
14	114.5	5.7	428	2	F85481	flavoprotein, elec
15	114.5	5.7	504	2	G87532	tryptophan halogen
16	113.5	5.6	428	2	C64725	fixC protein - Esc
17	110.5	5.5	405	2	T37022	probable lycopene
18	108.5	5.4	3161	2	T30342	protein Hmw1 - Ye
19	106.5	5.3	410	2	D75475	lycopene cyclase -
20	105	5.2	398	2	D83122	probable FAD-depen
21	105	5.2	557	2	F83181	probable GMC-type
22	104.5	5.2	402	2	AC3797	2-octaprenyl-6-met
23	104.5	5.2	402	2	C97576	2-octaprenyl-6-met
24	104	5.1	413	2	D86920	probable UDP-galac
25	103.5	5.1	722	2	B75608	GMC oxidoreductase
26	103.5	5.1	3163	2	T17440	probable polyketid
27	103.5	5.1	3163	2	AG0233	yersiniabactin bio
28	102.5	5.1	392	2	AB0873	2-octaprenyl-6-met
29	101.5	5.0	391	1	D64801	probable monooxyge

ALIGNMENTS

RESULT 1

S52585

dycopen cyclase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Dec-1999

C:Accession: S52585

R:Lin, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52585

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-382 <LIN>

A:Cross-references: EMBL:M06098; NID:G148393; PIDN:AAA21262.1; PID:G148396

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: Erwinia dycopen cyclase

Query Match 84.9%; Score 1715; DB 2; Length 382;

Best Local Similarity 83.8%; Pred. No. 1.6e-135;

Matches 320; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

Qy	1	MQPHYDLILVAGLANGLIALRLQQQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHR	60
Db	1	MRPHYDLILVAGLANGLIALRLQQQHPDMRILLIEAGPEAGGNHTWSFHDADLTESQHR	60
Qy	61	WIAPLVVHWPDYQVFPQRRHVNSGYCVTSRPHAGILRQFGQHLWLHTAVSAVHAE	120
Db	61	WVAPLVVHWPDYQVFPFTRRRKNSGYFSVTSQRFPAEVLQKFGQHLWISRAVAEHAD	120
Qy	121	SVQLADGRTHASTVIDGRGYPDSALRVGFOAFIQEWQLSAPHLSSPIIMDATVQQ	180
Db	121	AVLANNQVISAASAVIDGRGYTPNSALNVGFOAFIQEWRLSKPHGLSSPIIMDATVQQ	180
Qy	181	NGYRFVYIPLSATALLIEDTHYIDKANLQARRAQNRIDYAAAGWPLQTLRBEQCAL	240
Db	181	NGYRFYISFLSATALLIEDTHYIDNATLEPERAQNRIDYAAAGWPLQTLRBEQCAL	240
Qy	241	PIITLGDNRQFWQQQPOACSGLRAGLFFPTTCYSUPLAVALADRLSALDVFTSSVHTI	300
Db	241	PIITLGDSTAFWQQQPLACSGLRAGLFFPTTCYSUPLAVALADRLSALDVFTSSVHTI	300
Qy	301	AHPAQQRWQQQGFPMNLNMLFLAGPARRVNRVQRVYGLPEDLIARFYAGKLTVDRLR	360
Db	301	THFAHERWQQQGFPMNLNMLFLAGPARRVNRVQRVYGLPEDLIARFYAGKLTVDRLR	360
Qy	361	ILSGKPPVPVFAALQAIMTTHR	382
Db	361	ILSGKPPVPVLAALQAIMTTHR	382

RESULT 2

C37802
crry protein - Erwinia uredovora
C:Species: Erwinia uredovora
C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
C:Accession: C37802
R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim
J. Bacteriol. 172, 6704-6712, 1990
A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio
A:Reference number: A37802; MUID:91072214; PMID:2254247
A:Accession: C37802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <MIS>
A:Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4126.1; PID:g216684
A:Superfamily: Erwinia dycopene cyclase

Query Match 83.0%; Score 1678; DB 2; Length 382;
Best Local Similarity 82.2%; Pred. No. 1.9e-132;
Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

2Y 1 MOPHYDLILVGGGLANGLIARLQOQHEDMEILLIIEAGPEAGNHTWSFHEEDLTNLQHR 60
DB 1 MOPHYDLILVGGGLANGLIARLQOQHEDMEILLIIEAGPEAGNHTWSFHEEDLTNLQHR 60
2Y 61 MIAPLVVHHWPDYQVRRFRRHNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
DB 61 MIAPLVVHHWPDYQVRRFRRHNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
2Y 121 SVQLADGRIIHASTVIDGRGYTPDSALRVGFOAGIIGOEKWSLAPHSPIINDATVDQ 180
DB 121 SVRLKKGQVIGARVIDGRGYAANSALSGVFOAGIIGOEKWSLAPHSPIINDATVDQ 180
2Y 181 NGYRFTVTLPLSATALLEDTHYIDKMLQARARQNIIDYAAQGWPLQTLRLREEGAL 240
DB 181 NGYRFTVTLPLSATALLEDTHYIDKMLQARARQNIIDYAAQGWPLQTLRLREEGAL 240
2Y 241 PITLTDNRQWQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTQI 300
DB 241 PITLTDNRQWQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTQI 300
2Y 301 AHFAQRQOQOQGFRLNMLFLAGPAESRVRVMQRFYGLPDLIARFYAGKLTVTDLRL 360
DB 301 THFAERWQOQOQGFRLNMLFLAGPAESRVRVMQRFYGLPDLIARFYAGKLTVTDLRL 360
2Y 361 ILSGKPPVPVPAALQAIMTTHR 382
DB 361 ILSGKPPVPVPAALQAIMTTHR 382

RESULT 3
S22981
lycopen cyclase - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C:Accession: S52981
R: Hundle, B.; Alberti, M.; Niveltstein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Bur
Mol. Gen. Genet. 245, 406-416, 1994
A:Title: Functional assignment of Erwinia herbicola Eho 10 carotenoid genes expressed in
A:Reference number: S52976; MUID:95107236; PMID:7808389
A:Accession: S52981
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <HUN>
A:Cross-references: ENBL:M87280; NID:g148404; PIDN:AAA64980.1; PID:g148411
A:Start codon: GTG
A:Superfamily: Erwinia dycopene cyclase

Query Match 56.8%; Score 1148.5; DB 2; Length 386;
Best Local Similarity 58.3%; Pred. No. 3.3e-88;
Matches 218; Conservative 53; Mismatches 102; Indels 1; Gaps 1;

2Y 6 DLIIVGGLANGLIARLQOQHEDMEILLIIEAGPEAGNHTWSFHEEDLTNLQHRWIAPL 65

DB 3 DLIIVGGLANGLIARLQOQHEDMEILLIIEAGPEAGNHTWSFHEEDLTNLQHRWIAPL 62
QY 66 VVHHWPDYQVRRFRRHNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAEVOLA 125
DB 63 VAHAMFGYEVQPPDLRRRLAGYSITSERFAEALHQAALGENIWLNCSEVLPNSVRLA 122
QY 126 DGRITHASTVIDGRGYTPDSALRVGFOAGIIGOEKWSLAPHSPIINDATVDQOQYRF 185
DB 123 NGEALLAGAVIDGRGYTASSAMQTYQLFQOQWRLTQPHGLTVPILMDATVAQOQYRF 182
QY 186 VYTLPLSATALLEDTHYIDKMLQARARQNIIDYAAQGWPLQTLRLREEGALPITLT 245
DB 183 VYTLPLSATALLEDTHYIDKMLQARARQNIIDYAAQGWPLQTLRLREEGALPITLT 242
QY 246 GDNRFQWQOQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTQIAHFA 304
DB 243 GDIQALWADAPGVPRSGMRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTQIAHFA 302
QY 305 QORWQOQOQOQGFRLNMLFLAGPAESRVRVMQRFYGLPDLIARFYAGKLTVTDLRLTSG 364
DB 303 EXHMRQOQOQGFRLNMLFLAGPAESRVRVMQRFYGLPDLIARFYAGKLTVTDLRLTSG 362
QY 365 KPPVPVPAALQAIM 378
DB 363 KPPVPVPAALQAIM 376

RESULT 4
T07082
lycopen epsilon-cyclase (EC 5.5.1.1) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2003
C:Accession: T07082
R: Ronen, G.; Cohen, M.; Zamir, D.; Hirschberg, J.
submitted to the EMBL Data Library, July 1997
A:Description: Regulation of expression of the gene for lycopen epsilon cyclase during
A:Reference number: Z15905
A:Accession: T07082
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-526 <RON>
A:Cross-references: ENBL:Y14387; NID:e1285211; PIDN:CAA74745.1; PID:e1285212
A:Experimental source: cultivar VF36; leaf
C:Genetics:
A:Gene: CrTL-e-1
A:Map position: 12
C:Function:
A:Description: converts lycopen to delta-carotene
C:Superfamily: tomato lycopen beta-cyclase
C:Keywords: intramolecular lyase; isomerase

Query Match 7.5%; Score 152; DB 2; Length 526;
Best Local Similarity 21.1%; Pred. No. 8.9e-05;
Matches 87; Conservative 66; Mismatches 167; Indels 92; Gaps 17;

QY 6 DLIIVGGLANGLIARLQOQHEDMEILLIIEAGPEAGNHTWSFHEEDLTNLQHRWIA 62
DB 107 DLWVGCGPAG--LALAAEASAKULAGLVG--GPDLPFTNNYGVWDEDFKDLGLQ----- 157
QY 63 APLVHHWPDYQVRRFRRHNSGYCVTSRHF-----AGILRQOQOHLWHL 111
DB 158 ACIEHWRTDITVYLDDBDEPILIGRVRVSRHFLHEELKRCVEAGV-----LYLN 208
QY 112 T-----AVSAVHAEVQLADGRII---HASTVIDGRG-----YTPDSALRVGFOAGI 158
DB 209 SKVDRIEATNGQSLVECEGDWVPCFVTVVAGSAAGKFLQYELGSP-RVSQYTAIGVE 267
QY 159 WQL-SAPHGSLSSPIINDATVDQOQYR-----FVYTLPLSATALLEDT 201
DB 268 VEVDNNPFDLSLVFMD-----YDYLRLHDAQSLAEKYPFLYAMFMSPTVPEET 319
QY 202 HYDKANLQARARQNIIDYAAQGWPLQTLRLREEGALPITLTGDNRFQWQOQOQCSG 261

;Status: preliminary; nucleic acid sequence not shown.

;Molecule type: mRNA

;Residues: 1-500 <PEC>

;Cross-references: EMBL:X86452

;Superfamily: tomato lycopene beta-cyclase

;Keywords: intramolecular lyase; isomerase

Query Match 6.6%; Score 133; DB 2; Length 500;
Best Local Similarity 19.4%; Pred. No. 0.0032;
Matches 80; Conservative 74; Mismatches 152; Indels 106; Gaps 19;

Y 6 DLIVGAGLANGLIARLQQQHPDKVILLIEAGPEA--GGNH-----45

b 85 DLAVVGGGPAG--LAVAGQVSEAGLSVCSIDNPXKLIWPNVYGVVWDEFEAMDLLDCLDA 142

Y 45 TWS-----FHEEDITLQNHRIAPLVVHHWPDYQVFPQRRRHVNSGYCVTSRHFAGILR 101

b 143 TWSGAAYIDDNTAKDLR-----PYGRVNRKQKSKMMQ--KCIWN-----182

Y 102 QDFGQHLWLHTAVSAVHABSVOQA--DGRITASTVIDGRGYTP-----DSALRVGPOA 153

b 183 ---GVKHFQAKVIVHEBSKMLICNDGITIQATVVLDTGFSRSLVQYDKPYNPGVQV 239

Y 154 FICQEWQLSA-PHGLSSPIIMD-----ATVQQNGY--RFVITPLSATALLIEDT 201

b 240 AYGLIAEVEEHPFVNKMVFMWDRSHLKNNTDLKERNRIPTFLYAMPFSSNRIFLEET 299

Y 202 HYIDKANQAERARQIRDYAARQWQPLQTLRERQ-----GALPITLTDGNDQFQWQQ 255

b 300 SILVAPGLRIDDIQERVARLNLHGIVKVSIEDEHCLIPMGGLPVL-----347

Y 256 PQACSGL--RAGLFTPTTGYSLPLAVALADRL--SALDVTSSVHQTIAHFAQQRW-----308

b 348 PQRVVIGGTAGVHPSTGPMVARTLAAAPVAVANAIQVLSGSRSHSGNELSTAVWKDLW 407

Y 309 -----QQQGFERM-LNRMFLAGPESRWVQRYGVLPEDLIARFYAGKIT 354

b 408 PIERRQREFFCGMDILKDLPA-----TRFFDAFFDLEPRYMHGFLS 453

RESULT 11

70888

robable glf protein - Mycobacterium tuberculosis (strain H37RV)

;Species: Mycobacterium tuberculosis

;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

;Accession: E70888

;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.

;Andream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

;Authors: Squares, R.; Skelton, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

;Reference number: A70500; MUID:98295987; PMID:9634230

;Accession: E70888

;Status: preliminary; nucleic acid sequence not shown; translation not shown

;Molecule type: DNA

;Residues: 1-399 <COL>

;Cross-references: GB:AL022076; GB:AL1123456; NID:g3256026; PIDN:CAAL17873.1; PID:g295042

;Experimental source: strain H37RV

;Genetics:

;Gene: glf

;Superfamily: Methanobacterium thermoautotrophicum UDP-galactopyranose mutase

Query Match 6.3%; Score 127.5; DB 1; Length 399;

Best Local Similarity 27.6%; Pred. No. 0.0068;

Matches 83; Conservative 22; Mismatches 91; Indels 105; Gaps 18;

Y 1 MQPMTARFDLFWVGSGFFGLNIAERVAQ--LDKRVLVLERPHIGGN-----46

b 1 MQPMTARFDLFWVGSGFFGLNIAERVAQ--LDKRVLVLERPHIGGN-----46

Y 58 QERNAPLVVHHWPDYQVRFPPRRHNSGYCVTSRHFAGILQQQCGQHLWLHTAVSAV 117

b 58 QERNAPLVVHHWPDYQVRFPPRRHNSGYCVTSRHFAGILQQQCGQHLWLHTAVSAV 117

Db 47 -----AYSEABPO-----TGIEVHKYGAHLF-HTSKRV 74

QY 118 HABSVOQLADGRITIIHASTVIDGRGYTPDSALRVGPOAFCQAEQWQLSAPHGLSSPIIMDATV 177

Db 75 WDYVRQFTD-----FT-DYRHRV-FAMENQAVQF--PMGL-----GLV 109

QY 178 DQONGYRFVYVITPLSATALLIEDTHVIDKANLQ--AERA-----QNIRDYAAARQ 226

Db 110 SQPFG--KYFTEQARQIARQAAIDTADQNLEKALISLGRPLRYFAFVKGYTAKQ- 165

QY 227 WPLQTLREEQGA---LPITLTGDNQRFQWQQPOACSGLRAGLFPHFTTGYSLPLAVALA 282

Db 166 W--QTDPKELPAANITRLPVRVTDNRVF-----SDTYEGL--PTDGYTAMLQNMMA 213

QY 283 D 283

Db 214 D 214

RESULT 12

AD0584

Probable monooxygenase yleB [imported] - Salmonella enterica subsp. enterica serovar Typh

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C;Accession: AD0584

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moutie, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AD0584

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-391 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD05142.1; PID:g16501915; GSPDB:GN00176

C;Genetics:

A;Gene: yleB

C;Superfamily: ubiH protein

Query Match 5.8%; Score 117.5; DB 2; Length 391;

Best Local Similarity 23.4%; Pred. No. 0.045;

Matches 97; Conservative 42; Mismatches 163; Indels 113; Gaps 20;

QY 6 DLIVGAGLANGLIARLQQQ-----HPDMRILLIEAGPEA--CGN 44

Db 7 ELAIVGGVGGGALALGLAQGGFTVMVIEHAPAPFVADSQPDVRIISAASVALLKGL 66

QY 45 HTWSFHEEDLTINQHRWTAPLVVHHWPDYQVRFPPRRHNSGYCVTSRHFAGILRQOF 104

Db 67 GYW---EAVQGNRSHPY--RRLFTWENAHVFDAAELKPLLGVMVEN-----NVLOAL 118

QY 105 GQHLMLHTAV-----SAVHAE-----SVQLADGRITIIHASTVI--DGRGYTPDSALRVG 150

Db 119 WQALAHAPGVTLRVFASLAALHRRHDGYALEADLGEWTPKLVIGADGANSQVRQMGAG 178

QY 151 FOAFITGOEWQLSAPHGLSSPIIMDATVDQONG-----YRFVYITPLSATALLIEDTHYI 204

Db 179 IHA-----WQYAQ-----CMLITVKENAPGDSWQQTPTGPRAPFLP-----FD 220

QY 205 DKANLOAERARQNTIDYARQWQPLQTLRRE-----QGALPITLTGDNQRFW 252

Db 221 DWASLVWYDAPARINQLOSLMTQIQVEINQHFHARGVNVFVAGAPFLT-----RRHA 275

QY 253 QQQPO---ACSELGAGLFHPTTGYSLPLAVALADRLSALDVTSSVH--QTIAHFAQOR 307

Db 276 LQYAPGLALVGDAAHTHPLAGQGVNLGYRVDAL--IDVLASARSVGESWASHSVLR 333

QY 308 WQQQGFERMMLNMLEFLAGPESRWVQRYGVLPEDLIARFYAGKITVTDRLRL 362

Db 334 YQTR---SMADNFMQSG-----MDL---FYAGFSNELPPLRL 366

71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

March completed: February 29, 2004, 14:52:50
Job time : 15.6328 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:51:24 ; Search time 25.3626 Seconds
(without alignments)
3190.293 Million cell updates/sec

itle: US-09-941-947A-30

fect score: 2021

quence: 1 MOPHYDLIVGAGLIANGLIA.....SGKEPVPVFARLQAIMTTHR 382

oring table: BLOSUM62

atched: 809742 segs, 21153259 residues

atched: Gapop 10.0 , Gapext 0.5

al number of hits satisfying chosen parameters: 809742

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	2021	100.0	382	10	US-09-941-947A-30	Sequence 30, Appl
2	2021	100.0	382	14	US-10-218-118-6	Sequence 6, Appl
3	783.5	38.8	382	9	US-09-547-267-7	Sequence 7, Appl
4	783.5	38.8	382	10	US-09-920-923-5	Sequence 5, Appl
5	154.5	7.6	376	14	US-10-128-713A-20	Sequence 20, Appl
6	152	7.5	526	9	US-09-323-998D-53	Sequence 53, Appl
7	148.5	7.3	501	9	US-09-323-998D-55	Sequence 55, Appl
8	142.5	7.1	502	9	US-09-323-998D-56	Sequence 56, Appl
9	141.5	7.0	503	9	US-09-323-998D-61	Sequence 61, Appl
10	140	6.9	500	9	US-09-323-998D-59	Sequence 59, Appl
11	139.5	6.9	511	9	US-09-323-998D-60	Sequence 60, Appl
12	139	6.9	498	9	US-09-323-998D-57	Sequence 57, Appl
13	139	6.9	530	15	US-10-335-846-6	Sequence 6, Appl
14	134.5	6.7	516	9	US-09-323-998D-54	Sequence 54, Appl
15	133	6.6	500	9	US-09-323-998D-58	Sequence 58, Appl

Sequence 20, Appl
Sequence 51, Appl
Sequence 5, Appl
Sequence 23, Appl
Sequence 50, Appl
Sequence 47, Appl
Sequence 26, Appl
Sequence 3, Appl
Sequence 12332, A
Sequence 2, Appl
Sequence 21, Appl
Sequence 49, Appl
Sequence 8560, Ap
Sequence 12003, A
Sequence 52, Appl
Sequence 2, Appl
Sequence 25, Appl
Sequence 48, Appl
Sequence 13043, A
Sequence 8840, Ap
Sequence 27, Appl
Sequence 4489, Ap
Sequence 14057, A
Sequence 8419, Ap
Sequence 7248, Ap
Sequence 4784, Ap
Sequence 7544, Ap
Sequence 60, Appl
Sequence 12364, A
Sequence 21815, A

16 132.5 6.6 456 9 US-09-323-998D-20
17 130.5 6.5 529 9 US-09-323-998D-51
18 130.5 6.5 529 15 US-10-335-846-5
19 129.5 6.4 529 9 US-09-323-998D-23
20 129.5 6.4 529 9 US-09-323-998D-50
21 127.5 6.3 529 9 US-09-323-998D-47
22 127 6.3 533 9 US-09-323-998D-26
23 125.5 6.2 534 15 US-10-335-846-3
24 122.5 6.1 520 15 US-10-369-493-12232
25 122.5 6.1 524 9 US-09-323-998D-2
26 122.5 6.1 524 9 US-09-323-998D-21
27 122.5 6.1 524 9 US-09-323-998D-49
28 119.5 5.9 397 14 US-10-156-761-8560
29 115 5.7 436 15 US-10-369-493-12003
30 115 5.7 533 9 US-09-323-998D-52
31 115 5.7 533 15 US-10-335-846-2
32 113.5 5.6 378 9 US-09-323-998D-25
33 113.5 5.6 378 9 US-09-323-998D-48
34 112 5.5 529 15 US-10-369-493-13043
35 112 5.5 529 15 US-10-369-493-8840
36 109.5 5.4 374 9 US-09-323-998D-27
37 108.5 5.4 537 15 US-10-369-493-4489
38 107 5.3 538 15 US-10-369-493-14057
39 106.5 5.3 374 14 US-10-156-761-8419
40 106.5 5.3 529 15 US-10-369-493-7248
41 103.5 5.1 524 15 US-10-369-493-4784
42 103.5 5.1 524 15 US-10-369-493-7544
43 102.5 5.1 933 13 US-10-001-189-60
44 101 5.0 426 15 US-10-369-493-12364
45 100 4.9 528 15 US-10-369-493-21815

ALIGNMENTS

RESULT 1

US-09-941-947A-30
; Sequence 30, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 30
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-30

Query Match 100.0%; Score 2021; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.9e+199;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOPHYDLIVGAGLIANGLIARLQQQPDNRILLIIRAGPAGGNTWSPHEEDLTINQHR 60
Db 1 MOPHYDLIVGAGLIANGLIARLQQQPDNRILLIIRAGPAGGNTWSPHEEDLTINQHR 60
Qy 61 WIAFLVHHPDYQVFPQRRRHVNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120

Db 61 WIAELVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
2y 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
Db 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
2y 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGAL 240
Db 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGAL 240
2y 241 PITITGDNROFWQOQOACSLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSSVHQTII 300
Db 241 PITITGDNROFWQOQOACSLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSSVHQTII 300
2y 301 AHFAQQRWQOQOQFFRMLNRLFLAGPAESRWVRWQRYGLPEDLIARFYAGKLTVTDRLR 360
Db 301 AHFAQQRWQOQOQFFRMLNRLFLAGPAESRWVRWQRYGLPEDLIARFYAGKLTVTDRLR 360
2y 361 ILSGKPPVPVFAALQAIMTTHR 382
Db 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 2

JS-10-218-118-6
; Sequence 6, Application US/10218118
; Publication No. US20030148319A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia
; APPLICANT: Rouviere, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
; FILE REFERENCE: CL1876 US NA
; CURRENT APPLICATION NUMBER: US/10/218,118
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 60/312,646
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: Pantoea stewartii
JS-10-218-118-6

Query Match 100.0%; Score 2021; DB 14; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.9e-199;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 MQPHYDLIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHR 60
Db 1 MQPHYDLIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHR 60
2y 61 WIAELVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
Db 61 WIAELVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
2y 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
Db 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
2y 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGAL 240
Db 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGAL 240
2y 241 PITITGDNROFWQOQOACSLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSSVHQTII 300
Db 241 PITITGDNROFWQOQOACSLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSSVHQTII 300
2y 301 AHFAQQRWQOQOQFFRMLNRLFLAGPAESRWVRWQRYGLPEDLIARFYAGKLTVTDRLR 360

Db 301 AHFAQQRWQOQOQFFRMLNRLFLAGPAESRWVRWQRYGLPEDLIARFYAGKLTVTDRLR 360
Qy 361 ILSGKPPVPVFAALQAIMTTHR 382
Db 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 3

US-09-547-267-7
; Sequence 7, Application US/09547267
; Patent No. US20020147371A1
; GENERAL INFORMATION:
; APPLICANT: Rohmann, Hans-Peter
; APPLICANT: Passantes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/547,267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,645
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pokras, Bruce A.
; REGISTRATION NUMBER: 32,748
; REFERENCE/DOCKET NUMBER: RAN 6002/170
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-5801
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-547-267-7

Query Match 38.8%; Score 783.5; DB 9; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.6e-72;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
Qy 5 YDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHRWIA- 63
Db 3 HDLLIAGAGLSGALLIARVDRPRDARIVMDARSGPSDQHTWSCHTDLS---PSWLAR 59
Qy 64 --PLVVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 121
Db 60 LSPIRRGWTDQEVAFDPHSHRRLLTTGYGSTEAGALIGLQ---GVDLRNHTVATDDTG 116
Qy 122 VQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 181
Db 117 ATLDGRIERACVIDARGAVETPLTVGFKVGVETIEDAPHGVERPMDATVPMD 176
Qy 182 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGALP 241
Db 177 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGWPLQTLREBQAGALP 235

242 ILLTGNRQFWQQQPOAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
236 IALAHDAIGFWRDHAQAVPVGLGAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 294
300 IAHFAQORWQQQGFPMNLNMLFLAGPAESRWVQRYFGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDRYRLQRFYLPOLIERFYAGRLTLADRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

Query Match 38.8%; Score 783.5; DB 10; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.6e-72;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

5 YDLIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGNHTWSPHEEDTLNQHRIAPL 63
3 HDLLIAGAGLSGALIALAVRRDPARIVMLDARSQSDQHTWSCHDTLS---PEWLAR 59

64 --PLVHWPDPYQVRPQRRHVSVCVTSRHFAGILRQOFGQHLMLHTAVSAHES 121
60 LSPIERGENTQEVAFDHSRLTGTGSIAGALIGLQ---GVDLRWHTVATLDDTG 116

122 VOLADRIIHAHVTDGRGYPDSALRVGFOAFICQEWQLSAPHGLSSPIIMDATVDQON 181
117 ATLTDGSRIEAACVIDARGAVETPHLTGVCQKFGVGEIETDAPHGVERPMIMDATVPQMD 176

182 GYRVYTLPLSATALLIEDTHVIDKANLQARARONIRDYAARQCHPLOTLLRBOGALP 241
177 GYRIYLLPFSFTRILLIEDTRYSGGDLDDGALAQSLDYAARGGTGQE-MRRERGILP 235

242 ILLTGNRQFWQQQPOAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
236 IALAHDAIGFWRDHAQAVPVGLGAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 294
300 IAHFAQORWQQQGFPMNLNMLFLAGPAESRWVQRYFGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDRYRLQRFYLPOLIERFYAGRLTLADRL 354

360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

RESULT 4
US-09-920-923-5
Sequence 5, Application US/09920923
Publication No. US20030022273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 382
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923-5

TITLE OF INVENTION: Genes Involved in Isoprenoid Compound Production
FILE REFERENCE: CL-1788
CURRENT APPLICATION NUMBER: US/10/128,713A
CURRENT FILING DATE: 2002-04-22
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Microsoft Office 97
SEQ ID NO 20
LENGTH: 376
TYPE: PRT
ORGANISM: Rhodococcus erythropolis
US-10-128-713A-20

Query Match 7.6%; Score 154.5; DB 14; Length 376;
Best Local Similarity 23.8%; Pred. No. 4.3e-07;
Matches 75; Conservative 46; Mismatches 119; Indels 75; Gaps 15;

6 DULIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGNHTWSPHEEDTLNQHRIAPL 65
9 DVIVVGGSPAGRALATRCIAR--QLTVVVDPHRVMVTPVSVWADLP---SML--- 59

66 VVHWPDPYQVRPQRRHV-NSG-----YVCVTSRHFAGILRQOFGQHLMLHTAV--- 114
60 -----PDEVIASRIERPVSWTSGQKTLDRIVCV-----LNTSLQSPLSHTSIKVRG 106

115 ---SAVHAESVQLADGRIIHAHVTDGRGYPDSALRVGFOAFICQEWQLSAPHGLSSPI 171
107 ERAQTLSTTTVCVDSGSLTGSVVVDAG--TDLAVTTAQTATG---MIVDRALADPI 160

172 I-----MDATVDOONGY-----RVVYTLPLSATALLIEDTHVIDKANLQARARONI 218
161 LGGSEAFMDWRTD--NGTSDATPSEFLYAVPLDDERVLEETCLVGRPAL-----GL 211

219 RDYAAARQWPLQTLIREBOGALPITLTGDNRFQWQQPOACS-----GLRAGLPHPTT 271
212 RELETR---LRTLRHRCGEVDPDAPVRVRVFAVEGRDSSPDGVLRFGGGLMHPGT 267

272 GYSLPLAVALADRLS 286
268 GYSVASSIAEADTWA 282

RESULT 6
US-09-323-998D-53
Sequence 53, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: SUN ZAIREN
APPLICANT: CUNNINGHAM JR., FRANCIS X.
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 53
LENGTH: 526
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-323-998D-53

Query Match 7.5%; Score 152; DB 9; Length 526;
Best Local Similarity 21.1%; Pred. No. 1.3e-06;
Matches 87; Conservative 66; Mismatches 167; Indels 92; Gaps 17;

242 ILLTGNRQFWQQQPOAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
236 IALAHDAIGFWRDHAQAVPVGLGAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 294
300 IAHFAQORWQQQGFPMNLNMLFLAGPAESRWVQRYFGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDRYRLQRFYLPOLIERFYAGRLTLADRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

Query Match 38.8%; Score 783.5; DB 10; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.6e-72;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

5 YDLIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGNHTWSPHEEDTLNQHRIAPL 63
3 HDLLIAGAGLSGALIALAVRRDPARIVMLDARSQSDQHTWSCHDTLS---PEWLAR 59

64 --PLVHWPDPYQVRPQRRHVSVCVTSRHFAGILRQOFGQHLMLHTAVSAHES 121
60 LSPIERGENTQEVAFDHSRLTGTGSIAGALIGLQ---GVDLRWHTVATLDDTG 116

122 VOLADRIIHAHVTDGRGYPDSALRVGFOAFICQEWQLSAPHGLSSPIIMDATVDQON 181
117 ATLTDGSRIEAACVIDARGAVETPHLTGVCQKFGVGEIETDAPHGVERPMIMDATVPQMD 176

182 GYRVYTLPLSATALLIEDTHVIDKANLQARARONIRDYAARQCHPLOTLLRBOGALP 241
177 GYRIYLLPFSFTRILLIEDTRYSGGDLDDGALAQSLDYAARGGTGQE-MRRERGILP 235

242 ILLTGNRQFWQQQPOAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
236 IALAHDAIGFWRDHAQAVPVGLGAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 294
300 IAHFAQORWQQQGFPMNLNMLFLAGPAESRWVQRYFGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDRYRLQRFYLPOLIERFYAGRLTLADRL 354

360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

RESULT 5
US-10-128-713A-20
Sequence 20, Application US/10128713A
Publication No. US20030170847A1
GENERAL INFORMATION:
APPLICANT: Bramucci, Michael G

QY 6 DLIIVGAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTWSPHEEDLTINQHRWI 62
DB 107 DLVWIGCGPAG--LALAAESAKLGLNVGLV--GPDLPFTNNYGVWEDFXDLGLQ-----157
QY 63 APLVHVHWPDYQVRPQRHRHNSGVYCVTSRHP-----AGLLRQOQFQOHLWLH 111
DB 158 -ACIEHAWEDTTIYLDDBEPILIGRAYGRVSRHFLHBECLKRCVEAGV-----LYLN 208
QY 112 T----AVSAVHAESVOLADGRII---HASTVIDGRG-----YTPDSALRVGFQAFIGOE 158
DB 209 SKVDRIEATNQSLVCEBGDVIPCRFVTVASGAASKFLQYELGSP-RVSVQYAVGVE 267
QY 159 WQL-SAPHGLSPLIMDATVDQONGY-----FYVITPLSATALLIEDT 201
DB 268 VEVDMNPFDPSSLMVMD-----YRDLRHAQSLKAPFTFLYAMPSPTRVFFFEET 319
QY 202 HYIDKANLOAERARQIRDYAARQGWPLQTLRLREEQ-----ALDVFTSSVH--QTIHFAQOQW 308
DB 320 CLASKDAMPFDLLKKXLMRLNTLGVRIKEIYEEMSVIPVCGSLN-----TEQKTLAFG 375
QY 262 LRAGLFHPTTGYSLPLAVALADRLSAL-----DVTSSSVHQTIAH-----PAQ 305
DB 376 AAASVHPATGYSVRSLSSEAPKCSVLANILRQHYXNMULTSSSIPSISTQAWNTLMPQ 435
QY 306 QWQOQGFPMENRMLFLAGPAESRWVQWQRYGYPGLPDLIARFYAGKLTWTD 357
DB 436 ERRKRSFF-LFGLALILQLDIEGIRSPFRAPRVYKMKWQOGLGSSLSAD 486

RESULT 7

US-09-323-998D-55
; Sequence 55, Application US/09323998D
; Patent No. US20020102631A1
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998D
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 08/937,155
; PRIOR FILING DATE: 1997-09-25
; PRIOR APPLICATION NUMBER: 08/624,125
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 55
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-323-998D-55

Query Match 7.3%; Score 148.5; DB 9; Length 501;
Best Local Similarity 21.0%; Pred. No. 2.7e-06;
Matches 92; Conservative 66; Mismatches 175; Indels 105; Gaps 20;

QY 6 DLIIVGAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTW--SFHEED-LTLNQH 59
DB 84 DLIAVGGGPAG--LAVAQVSEAGLSVCSIDPSPKLIWPNNYGVWVDFEAMDLLDCLDT 141
QY 60 RWIAPLVHWPDYQVRPQRHRHNSGVYCVTSRHFAGILRQOF-----GOHL 108
DB 142 TWSGAVV-----YDEGKQDLSEPYGRVNRKQLKSKMLOKCTINGVKF 185
QY 109 WLTAVSAVHAESVOLADGRII---HASTVIDGRGYP-----DSALRVGFQAFIGQEW 160
DB 186 HQSKVTNVVHEEANSWVCSGVKIQASVWLDATGFSRCLVQYDKPNPGVQVAYGVAE 245

QY 161 LSA-PHGLSSPIIMDATVDQONGY-----RFVYITPLSATALLIEDTHYIDKAN 208
DB 246 VDHHPDVPDKVFMWDWRKHLDSYPCLKERNSKIPTFLYAMPFSSNRIFLBEETSLVARPG 305
QY 209 LQAEARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG 262
DB 306 LRWEDIQERMAARLKHGLNVRIEEDERCVPIMGGPLVL-----PQVVG 353
QY 263 --RAGLFHPTTGYSLPLAVALADRLS-----ALDVFTSSVH--QTIHFAQOQW 308
DB 354 GGTAGWVHSTGYMARTLAAAPIVANAI VRYLGSPPSSNLRGDLSAEVWRDLWPIERR 413
QY 309 QOQOQGFPMENRMLFLAGPAESRWVQWQRYGYPGLP-----LIAREYACKLT 355
DB 414 ROREFF-CEGMDILKLDLDAATRRPFDAPFDLQPHYWGFLSSRLFLPELLVGLSLFSH 472
QY 356 ---TDRLRILSKPPVPV 370
DB 473 ASNTSRLEIMT-KGTVPL 489

RESULT 8

US-09-323-998D-56
; Sequence 56, Application US/09323998D
; Patent No. US20020102631A1
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998D
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 08/937,155
; PRIOR FILING DATE: 1997-09-25
; PRIOR APPLICATION NUMBER: 08/624,125
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 56
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Adonis palaeestina
US-09-323-998D-56

Query Match 7.1%; Score 142.5; DB 9; Length 502;
Best Local Similarity 21.9%; Pred. No. 1.1e-05;
Matches 70; Conservative 49; Mismatches 131; Indels 69; Gaps 15;

QY 6 DLIIVGAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTW--SFHEED-LTLNQH 59
DB 87 DLIAVGGGPAG--LAIQOQVSEAGLSVCSIDPSPKLIWPNNYGVWVDFEAMDLLDCLDT 144
QY 60 RWIAPLVHWPDYQVRPQR-----RRHNSGVY--CVTSRHFAGILRQOFGCHLWLH 111
DB 145 TWSGAVV--YTDNSKKYLDPRYGRVNRKQLKSKMLOKCVTN-----GVKPEQA 191
QY 112 TAVSAVHAESVOLADGRII---HASTVIDGRGYP-----DSALRVGFQAFIGQEWLSA 163
DB 192 KVKVIHEESKSLICNDGITINATVWLDATGFSRCLVQYDKPNPGVQVAYGVAEVEE 251
QY 164 -PHGLSSPIIMDATVDQONGY-----RFVYITPLSATALLIEDTHYIDKANLQA 211
DB 252 HPFDLKKLFWMDRSHLNKLEKDKNRKIPTFLYAMPFSSTKIFLEETSLVARPGLRF 311
QY 212 ERARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG--R 263
DB 312 EDIQERMAARLKHGLIKVKSIEEDERCVPIMGGPLVL-----PQVVGIGGT 359

Y 264 AGLEHFTTGYSLPLAVALA 282
||:||||:|:
b 360 AGWHPSTGYWVARTLAAA 378

RESULT 9

S-09-323-998D-61

Sequence 61, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

FILE OF INVENTION: METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998D

CURRENT FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 08/937,155

PRIOR FILING DATE: 1997-09-25

PRIOR APPLICATION NUMBER: 08/624,125

PRIOR FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 61

LENGTH: 503

TYPE: PRT

ORGANISM: Narcissus sp.

S-09-323-998D-61

Query Match 7.0%; Score 141.5; DB 9; Length 503;
Best Local Similarity 20.4%; Pred. No. 1.4e-05;
Matches 85; Conservative 66; Mismatches 150; Indels 115; Gaps 19;

Y 6 DLILVAG-----LANGIALRLQOQHPDMRILLIEAGPEAGGNHTWTFHEEDLTL 56
||:||||:|:
D 89 DLAVVGGGLARSCTSLGGGLSVSVID---PNPKLI----WPNYNG--VWVDFEDMDL 139
||:||||:|:
Y 57 ---NOHRWIAPIVHHWPDYQVRFPPQRERHNSGYCYVTSRHFAGILRQOP----- 104
||:||||:|:
D 140 LCLDATSGAIV-----YVDRSTKLSRPYARVNRKLSKMMKKCV 183
||:||||:|:
Y 105 ---GQHLWHTAVSAVHARS---VQLADGRITHASTVIDRGYTP-----DSALRVGQAP 154
||:||||:|:
D 184 SNGVRFHQATVYKAKHBEKSYLICSDDGVTIDARVYLDATGFSRLQVQYKYPNGYQA 243
||:||||:|:
Y 155 IQQEWQLSA-PHGLSSPIIMD-----ATVDOONGY--RFVYTLPLGATALLIEDTH 202
||:||||:|:
D 244 YGILAEVVEHPFDVDMKVPMDWRDHLNGKAEILNERNAKIPTFLYAMPFSSNRIPLETS 303
||:||||:|:
Y 203 YIDKANLQARONIRDYAARQGWPLQTLRREQ-----GALPITLTGDNROFMOQOP 256
||:||||:|:
D 304 LVARPLGKMDIQERWVARLNHLGIRIKSIEEDRCVTPMGGLPVI-----P 351
||:||||:|:
Y 257 QACSGI--RAGLFHPTTGYSLPLAVALADRLSALDVTSSVHTIAH----- 302
||:||||:|:
D 352 QRWVGIGTAGVHPSTGY-----MVARTLAAPIVANSIVQVLYSDSLGSLGNDLSADV 405
||:||||:|:
Y 303 -----FAOORWQQGFRRMLNRLFLAGPAERWRVQRYGLPDLIARFYAGKLT 354
||:||||:|:
D 406 WKDLNPIERRRQREFF-CFGMDILLKLDLTPA-----TERFFDAPFLEPRYWHGFLS 456
||:||||:|:

RESULT 10

S-09-323-998D-59

Sequence 59, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

Y TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 108172-09019

; CURRENT APPLICATION NUMBER: US/09/323,998D

; CURRENT FILING DATE: 1999-06-02

; PRIOR APPLICATION NUMBER: 09/088,724

; PRIOR FILING DATE: 1998-06-02

; PRIOR APPLICATION NUMBER: 09/088,725

; PRIOR FILING DATE: 1998-06-02

; PRIOR APPLICATION NUMBER: 08/937,155

; PRIOR FILING DATE: 1997-09-25

; PRIOR APPLICATION NUMBER: 08/624,125

; PRIOR FILING DATE: 1996-03-29

; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 59

; LENGTH: 500

; TYPE: PRT

; ORGANISM: Nicotiana sp.

US-09-323-998D-59

Query Match 6.9%; Score 140; DB 9; Length 500;
Best Local Similarity 20.2%; Pred. No. 2e-05;

Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;

QY 6 DLILVAGLANGIALRLQOQHPDMRILLIEAGPE---AGGNHTW--SFHEED-LTLNQH 59
||:||||:|:
DB 85 DLAVVGGGPAG--LAVAQGVSEAGISVVSIDPSPKLIWPNNYGVWVDFEAMDLDCILDA 142
||:||||:|:
QY 60 RWTAPLVHHWPDYQVRFPPQRERHNSGYCYVTSRHFAGILRQOPGQHLW-----L 110
||:||||:|:
DB 143 TWGTVV-----YIDNTTKDLDPYGRVNRKLSKMMQKILNGVKF 186
||:||||:|:
QY 111 HTA--VSVAHSAESVQLA---DGRITHASTVIDRGYTP-----DSALRVGQAPFGQBMQ 160
||:||||:|:
DB 187 HAAKVIKVIHESAKSMILCNDGVTIOATVLDATGFSRLQVQYKYPKPGYQVAYGILAE 246
||:||||:|:
QY 161 LSA-PHGLSSPIIMD-----ATVDOONGY--RFVYTLPLSATALLEDTHYIDKAN 208
||:||||:|:
DB 247 VEEHPFDTSKVLMDWRDHLNGKAEILNERNAKIPTFLYAMPFSSNKIPLETSUVRPG 306
||:||||:|:
QY 209 LQAEARONIRDYAARQGWPLQTLRREQ-----GALPITLTGDNROFMOQOPAC--S 260
||:||||:|:
DB 307 LRMDDIQERWVARLNHLGIRIKSIEEDRCVTPMGGLPVI-----PQVVGT 354
||:||||:|:
QY 261 GLRAGLFHPTTGYSLPLAVALADRLSALDVTSSSVH-----QTAHFAQQRW- 308
||:||||:|:
DB 355 GGTAGLVHPSTGY-----MVARTLAAAPVWANAIIHYLGSEKDLGNLSAAVWMDLWP 408
||:||||:|:
QY 309 -----QOQGFPRM-LNRLFLAGPAERWRVQRYGLPDLIARFYAGKLT 354
||:||||:|:
DB 409 IERRQRBEFFCFGMDILLKLDLPA-----TERFFDAPFLEPRYWHGFLS 453
||:||||:|:

RESULT 11

US-09-323-998D-60

Sequence 60, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

Y TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 108172-09019

; CURRENT APPLICATION NUMBER: US/09/323,998D

; CURRENT FILING DATE: 1999-06-02

; PRIOR APPLICATION NUMBER: 09/088,724

; PRIOR FILING DATE: 1998-06-02

; PRIOR APPLICATION NUMBER: 09/088,725

; PRIOR FILING DATE: 1998-06-02

; PRIOR APPLICATION NUMBER: 08/937,155

; PRIOR FILING DATE: 1997-09-25

; PRIOR APPLICATION NUMBER: 08/624,125

PRIOR FILING DATE: 1996-03-29
 NUMBER OF SEQ ID NOS: 61
 SOFTWARE: Patent in Ver. 2.1
 SEQ ID NO 60
 LENGTH: 511
 TYPE: PRT
 ORGANISM: Tagetes erecta
 IS-09-323-998D-60

Query Match 6.9%; Score 139.5; DB 9; Length 511;
 Best Local Similarity 21.0%; Pred. No. 2.4e-05;
 Matches 86; Conservative 57; Mismatches 133; Indels 59; Gaps 15;

6 DLILVAGLANGLIALRLQQHPDMRIILLIEAGPE---AGGNHTW---SFHEEDITLNOHR 60
 92 DLVAVGGSG--LAVAQVSEAGLTVCSIOIPSPKLIWPNYVWVDFEFAMDL-----143

61 WIAPLVVHHWPDYQVRFPOR--RRHNSGYCVTSRHFAGILRQOP---GQHLWLHTAVSA 116
 144 --LDCLDTTSSAVYIDEKSTKSLNRYARNRKQLTKMLQKCIANGVFKHQAKVIKV 201

117 VHAESVOLA---DGRITTHASTVIDGRGYTP-----DSALRVGFQAFIGQEWOLSA-PHGL 167
 202 IHEELKSLICNDGVTIQTATVLDATGFSRSIVQYDKFNPYQVAYGILAEVEEHPFDV 261

168 SSPILM---DATVDQNGYR-----FVYTLPLSATALLIEDTHYIDKANLQAEARQ 216
 262 DQMLFMDWEDSHLDQMLKAKNSRIPTFLYAMPFSSTRIFLETSLSVAPGLGMDIQE 321

217 NTRDYAARGNPLQTLRREQ-----GALPITITGDNRFQWQOQOACSGL--RAGLPH 268
 322 RMAYRLKHLGIVKYSIEDERCVIPMGGLPVL-----PQRLVIGIGTAGMVH 369

269 PTGYSLPLAVALA 282
 370 PSTGYVARTLAA 383

RESULT 12
 JS-09-323-998D-57
 Sequence 57, Application US/09323998D
 Patent No. US20020102631A1
 GENERAL INFORMATION:
 APPLICANT: CUNNINGHAM JR., FRANCIS X.
 APPLICANT: SUN, ZAIEN
 TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
 TITLE OF INVENTION: METHODS OF USE THEREOF
 FILE REFERENCE: 108172-09019
 CURRENT APPLICATION NUMBER: US/09/323,998D
 CURRENT FILING DATE: 1999-06-02
 PRIOR FILING DATE: 1999-06-02
 PRIOR APPLICATION NUMBER: 09/088,724
 PRIOR FILING DATE: 1999-06-02
 PRIOR APPLICATION NUMBER: 09/088,725
 PRIOR FILING DATE: 1999-06-02
 PRIOR APPLICATION NUMBER: 08/937,155
 PRIOR FILING DATE: 1997-09-25
 PRIOR APPLICATION NUMBER: 08/624,125
 PRIOR FILING DATE: 1996-03-29
 NUMBER OF SEQ ID NOS: 61
 SOFTWARE: Patent in Ver. 2.1
 SEQ ID NO 57
 LENGTH: 498
 TYPE: PRT
 ORGANISM: Capsicum sp.
 US-09-323-998D-57

Query Match 6.9%; Score 139; DB 9; Length 498;
 Best Local Similarity 20.4%; Pred. No. 2.6e-05;
 Matches 81; Conservative 70; Mismatches 170; Indels 76; Gaps 18;

6 DLILVAGLANGLIALRLQQHPDMRIILLIEAGPE---AGGNHTW---SFHEEDITLNOHR 60
 83 DLVAVGGSG--LAVAQVSEAGLSVCSIDPNPKLIWPNYVWVDFEFAMDL-----134

QY 61 WIAPLVVHHWPDYQVRFPOR--RRHNSGYCVTSRHFAGILRQOP---GQHLWLHTAVSA 116
 DB 135 --LDCLDATSGAAVYIDDKTTDLNRPYGRVNRKQLKSKMMQKCIANGVFKHQAKVIKV 192

QY 117 VHAESVOLA---DGRITTHASTVIDGRGYTP-----DSALRVGFQAFIGQEWOLSA-PHGL 167
 DB 193 IHEESKMLICNDGVTIQTATVLDATGFSRSIVQYDKFNPYQVAYGILAEVEEHPFDV 252

QY 168 SSPILM-----ATVDQNGYR--RFVYTLPLSATALLIEDTHYIDKANLQAEARQ 216
 DB 253 NKVYFMDWRDHLKNNVELKERSRIPTFLYAMPFSSTRIFLETSLSVAPGLGMDIQE 312

QY 217 NTRDYAARGNPLQTLRREQ-----GALPITITGDNRFQWQOQOACSGL--RAGLPH 268
 DB 313 RMVARLSHLGIVKYSIEDERCVIPMGGLPVL-----PQRLVIGIGTAGMVH 360

QY 269 PTGYSLPLAVALA--SALDVTSSVHOTTATPAQORW-----OOQGPFRM-L 317
 DB 361 PSTGYVARTLAAAPVYANAIQYLSERSHSGDELSAAVWKDLWPIERRRORPEFFCFQM 420

QY 318 NMCLFLAGPAESRWRVQRFYGLFEDLIARFYAGKLT 354
 DB 421 DILLKLDLPA-----TRRFFDAFFDLEPRYWHGFLS 451

RESULT 13
 US-10-335-846-6
 Sequence 6, Application US/10335846
 Publication No. US20030220405A1
 GENERAL INFORMATION:
 APPLICANT: CUNNINGHAM JR., FRANCIS X.
 APPLICANT: SUN, ZAIEN
 TITLE OF INVENTION: GENES ENCODING EPSILON LYCOPENE CYCLASE AND METHOD FOR
 TITLE OF INVENTION: PRODUCING BICYCLIC EPSILON CAROTENE
 FILE REFERENCE: 2747-0084-27 CIP
 CURRENT APPLICATION NUMBER: US/10/335,846
 CURRENT FILING DATE: 2003-01-03
 PRIOR APPLICATION NUMBER: US/09/084,222
 PRIOR FILING DATE: 1998-05-26
 PRIOR APPLICATION NUMBER: 08/937,155
 PRIOR FILING DATE: 1997-09-25
 PRIOR APPLICATION NUMBER: 08/624,125
 PRIOR FILING DATE: 1996-03-29
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: Patent in Ver. 2.0
 SEQ ID NO 6
 LENGTH: 530
 TYPE: PRT
 ORGANISM: Adonis palaeestina
 US-10-335-846-6

Query Match 6.9%; Score 139; DB 15; Length 530;
 Best Local Similarity 19.6%; Pred. No. 2.8e-05;
 Matches 79; Conservative 72; Mismatches 179; Indels 74; Gaps 15;

6 DLILVAGLANGLIALRLQQHPDMRIILLIEAGPE---AGGNHTW---SFHEEDITLNOHR 62
 109 DLVAVGGSG--LSLAEEAAKLGKVLGLI--GPDLPNTNNGVWDEFKOLGLR-----160

QY 63 APLVWHWPDYQVRFPORRRHNSGYCVTSRHFAGILRQOPGQHLWLHTAVSAVHAESV 122
 DB 161 --CIEHAWKDTIVYLDNDAPVLIGRAYGRVSRH---LHBEELKFC-VESGVSYLDSKVE 214

QY 123 OLADGRIIHAFTVIDGRGYTPDSALRVGFQAFIGQ---EWOLSAP-----HGL-----168
 DB 215 RITEAGDGHSLVCENEIPPCSLATVAGSAGSKLLEYVGGFRVCVOTAYGVEVEVEN 274

QY 169 SP-----IIMD-----ATVDQNGYRFPVYTLPLSATALLIEDTHYIDKANLQAEAR 214
 DB 275 NPYDENLWFMVDYRDMQKLCQCSBEETFLYVMPSPTRLPFEETCLASKDAMPFDLL 334

QY 215 RQNIIDYARQWPLQTLRBEQGLP.TLITGDNRFQWQOQOACSGLRLAGLPHPTTIGYS 274

335 KRKLMSRLKTLGIQVTKYBESWYIPVGSPLN-----TECKNLAGAASWVHRATGVS 390
275 LPLAVALADRL-----SALDVFTSSSVHQTIAHFAQORW-----QQQGF 313
391 VYRSEAPKYASVIATIKLDNQDNYVYVQSQA-----VNISMQWSSLPKXKRRQRAF 446
314 FEMLRMLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLTVD 357
447 PTEFGLLEIVOLDIATRTFTFRFTPLPTWMMWGLGSSLSFPD 490

SULT 14
-09-323-998D-54
Sequence 54, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIEN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 54
LENGTH: 516
TYPE: PRI
ORGANISM: Tagetes erecta
S-09-323-998D-54

Query Match 6.7%; Score 134.5; DB 9; Length 516;
Best Local Similarity 19.9%; Pred. No. 7.9e-05;
Matches 86; Conservative 65; Mismatches 188; Indels 93; Gaps 17;

Y 6 DLILVAGLGLALRLQOQHDPMDRILLTEAGPEAGKHVTSFEEDLTINQHWIAPL 65
b 96 DLVVGCGPAG--LALAGESAKGLINALL--GFDLPFTNNYGVWDEFI---GLGLEGC 148
Y 66 VHWHPDYQVPPORRRHNSGYVTSRHFAGILRQOFGHMLHTAVSAVHAESVOL- 124
b 149 IEHVWRDTV-----YLDNDPILIGRAYGRVSRDLLEELLTRCMESGVSYLSKVE 201
Y 125 -----ADGRI---IHASTVIDG--RGYTPDSAL---RVGFQARTIGQEWOL-S 162
b 202 RITEAPNGLSLIECGNITPFCRLATVAGSAAGSKLLQYELGPGPVCVOTAYGIEVEVS 261
Y 163 APHGLSSPIINDATVDOONGYR-----FVYTLPLSATALLIEDTHYIDK 206
b 262 TEYDPSLAVMD-----YRDYTKHSQSLEAQYPTFLYVWMSPTXVFFETCLASK 313
Y 207 ANLQAEARQNIIDYAAACQWPLQTLRLREQAGALPITLTGDNRCQWQOQPCASGLRAGL 266
b 314 EAMFPELLKTKLMSRLKWTGIRITKIVIEEWSYIPVGSPLN-----TECKNLAGAASW 369
Y 267 PHPTTGYSLPLAVALADRLSAL--DVFTSSSVHQTIAHFAQORW-----FAQORWQ 309
b 370 VHPATGYSVRSVSEAPNYAAVIAKILGKNSQKMLDHQRYTTNISKQAWETLPLKRXR 429
Y 310 QQGFPRMLNKLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLTVTORLILSGKPPVP 369
b 430 QRAFF-LFGLALIVQMDIEGTRTFTFRFTPLPTWMMWGLGSSLSSTDLI-----478
Y 370 VPAALQAIMTTH 381

Db 479 IPAFWFIAPH 490

RESULT 15
US-09-323-998D-58
Sequence 58, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIEN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 58
LENGTH: 500
TYPE: PRI
ORGANISM: Lycopersicon esculentum
US-09-323-998D-58

Query Match 6.6%; Score 133; DB 9; Length 500;
Best Local Similarity 19.4%; Pred. No. 0.00011;
Matches 80; Conservative 74; Mismatches 152; Indels 106; Gaps 19;

QY 6 DLILVAGLGLALRLQOQHDPMDRILLTEAGPEA--GGNH-----45
Db 85 DLAVVGGGPAG--LAVACQVSEAGLSVCSIDPRLKLIWNNYGVWVDEFEAMDLLDCLDA 142
QY 46 TMS-----FHEEDLTINQHWIAPLWVHEWPDYQVFPQRRHNSGYCVTSRHFAGILR 101
Db 143 TMSGAATVYIDDNTAKDLHR-----PYGRVNRKQLKSKMQ--KCMN-----182
QY 102 CQFGQHLWHTAVSAVHAESVOLA---DGRTHASTVIDRGVTP-----DSALRVGFOA 153
Db 183 ---GVKFRQAKVIKVIHEESKSLICNDGTITQATVLDATGFSRSLVQYDXYNFGYV 239
QY 154 FIGQEWOLSA-PLGLSSPIIND-----ATVDOONGY--RFVYTLPLSATALLIEDT 201
Db 240 AVGILARVEEHPFDVNNVFMWRDLSHLKNTDLKXNSRIPTFLYAMPFSSNRIPLBET 299
QY 202 HYIDKANQAERARQNIIDYAAACQWPLQTLRLREQ-----GALPITLTGDNRCQWQO 255
Db 300 SIVARPGLRIDDIQERVARLNHLGIKVKSIIEDEHCLIPMGSPPLVL-----347
QY 256 PQACSGL--RAGLPHPTTGYSLPLAVALADRL--SALDVFTSSSVHQTIAHFAQORW-----308
Db 348 FORVVIGGTAGWHPSTGYVARTAAAEVANAIIQYLGSRSHSGNELSTAVWKOLW 407
QY 309 -----QQQGFPRM-LNKLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLT 354
Db 408 PIERERQREFFCFGMDILLKDLPA-----TRRFTDAFFDLEPRVWHGFLS 453

Search completed: February 29, 2004, 15:28:00
Job time : 27.3626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:35:44 ; Search time 11.6083 Seconds
(without alignments)
1698.885 Million cell updates/sec

itle: US-09-941-947A-30

erfect score: 2021
equence: 1 MDPHYDLILVAGLANGLIA.....SGKPPVPFPAALQIMTHR 382

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

ral number of hits satisfying chosen parameters: 389414

imum DB seq length: 0
iximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :
Issued Patents AA:*
1: /cgn2_6/ptodata/2/aa/5A.COMB.pdp:*
2: /cgn2_6/ptodata/2/aa/5B.COMB.pdp:*
3: /cgn2_6/ptodata/2/aa/6A.COMB.pdp:*
4: /cgn2_6/ptodata/2/aa/6B.COMB.pdp:*
5: /cgn2_6/ptodata/2/aa/ECTUS.COMB.pdp:*
6: /cgn2_6/ptodata/2/aa/backfiles.pdp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1678	83.0	382	1 US-07-783-705A-3	Sequence 3, Appli
2	1127	55.8	374	1 US-08-095-726-14	Sequence 14, Appl
3	1127	55.8	374	1 US-08-096-623A-14	Sequence 14, Appl
4	783.5	38.8	382	3 US-08-660-645A-7	Sequence 7, Appli
5	783.5	38.8	382	3 US-09-298-718-7	Sequence 7, Appli
6	783.5	38.8	382	3 US-09-546-969-7	Sequence 7, Appli
7	783.5	38.8	382	3 US-08-980-832-5	Sequence 5, Appli
8	783.5	38.8	382	4 US-09-547-267-7	Sequence 7, Appli
9	783.5	38.8	382	4 US-09-920-923B-5	Sequence 5, Appli
10	759.5	37.6	386	2 US-08-663-310-6	Sequence 6, Appli
11	759.5	37.6	386	2 US-09-008-491-6	Sequence 6, Appli
12	759.5	37.6	386	3 US-09-335-919-6	Sequence 6, Appli
13	154.5	7.6	498	3 US-09-134-607A-19	Sequence 19, Appl
14	152.5	7.5	498	3 US-09-134-607A-17	Sequence 17, Appl
15	152.5	7.5	498	3 US-09-134-607A-18	Sequence 18, Appl
16	152	7.5	526	4 US-09-323-998E-53	Sequence 53, Appl
17	148.5	7.3	411	1 US-08-399-561-2	Sequence 2, Appli
18	148.5	7.3	498	2 US-08-702-598-2	Sequence 2, Appli
19	148.5	7.3	501	4 US-09-323-998E-55	Sequence 55, Appl
20	142.5	7.1	502	4 US-09-323-998E-56	Sequence 56, Appl
21	141.5	7.0	503	4 US-09-323-998E-61	Sequence 61, Appl
22	140	6.9	500	4 US-09-323-998E-59	Sequence 59, Appl
23	139.5	6.9	511	3 US-09-201-641-2	Sequence 2, Appli
24	139.5	6.9	511	4 US-09-323-998E-60	Sequence 60, Appl
25	139	6.9	498	4 US-09-323-998E-57	Sequence 57, Appl
26	134.5	6.7	516	3 US-09-201-641-6	Sequence 6, Appli
27	134.5	6.7	516	4 US-09-323-998E-54	Sequence 54, Appl

Query Match 83.0%; Score 1678; DB 1; Length 382;

Sequence 58, Appl
Sequence 20, Appl
Sequence 20, Appl
Sequence 51, Appl
Sequence 23, Appl
Sequence 47, Appl
Sequence 50, Appl
Sequence 26, Appl
Sequence 2, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 49, Appl
Sequence 21, Appl
Sequence 6479, Ap
Sequence 52, Appl
Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-07-783-705A-3
; Sequence 3, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeyuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladass & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783,705A
; FILING DATE: 19911023
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-103078
; FILING DATE: 21-APR-1989
; APPLICATION NUMBER: JP 2-53225
; FILING DATE: 05-MAR-1990
; APPLICATION NUMBER: US 07/519,011
; FILING DATE: 19-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwadron, Janet I.
; REGISTRATION NUMBER: 33,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-708-1935
; TELEFAX: 212-246-5959
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-705A-3

Best Local Similarity 82.2%; Pred. No. 1.2e-159;
Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

QY 1 MQPHYDLILVAGLANGLIALRLOQQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRIAPL 60
DB 1 MQPHYDLILVAGLANGLIALRLOQQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRIAPL 60
QY 61 WTAPLVVHHWPDYQVRFQRRRHVNSGYCVTSRHFAGILRQOQGHLMHTAVSAVHAZ 120
DB 61 WTAPLVVHHWPDYQVRFQRRRHVNSGYCVTSRHFAGILRQOQGHLMHTAVSAVHAZ 120
QY 121 SVQLADGRIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDQ 180
DB 121 SVQLADGRIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDQ 180
QY 181 NGYRFVYVTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGNPLOTLLREEGAL 240
DB 181 NGYRFVYVTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGNPLOTLLREEGAL 240
QY 241 PITLGDNRQFQQOPOACSGLRAGLFHPTTGYSLPLAVADRLSALDVFTSSVHCTI 300
DB 241 PITLGDNRQFQQOPOACSGLRAGLFHPTTGYSLPLAVADRLSALDVFTSSVHCTI 300
QY 301 ASHPAQRWQOQGFPMNLNMLFLAGPAESRWVQRYGFLPDLIARFYAGKLTVDRLR 360
DB 301 ASHPAQRWQOQGFPMNLNMLFLAGPAESRWVQRYGFLPDLIARFYAGKLTVDRLR 360
QY 361 ILSGKPPVVFVFAALQAIMTTHR 382
DB 361 ILSGKPPVVFVFAALQAIMTTHR 382

RESULT 2
US-08-095-726-14
; Sequence 14, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Huel-Che B.
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,726
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/785,566
; FILING DATE: 30-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5530188val B
; TELEPHONE: 3128567180
; TELEFAX: 3128564972
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-095-726-14

Query Match 55.8%; Score 1127; DB 1; Length 374;
Best Local Similarity 58.4%; Pred. No. 1.7e-104;
Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

QY 6 DLILVAGLANGLIALRLOQQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRIAPL 65
DB 3 DLILVAGLANGLIALRLOQQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRIAPL 62
QY 66 VVHWPDYQVRRRHVNSGYCVTSRHFAGILRQOQGHLMHTAVSAVHAZ 125
DB 63 VAHWPDYQVRRRHVNSGYCVTSRHFAGILRQOQGHLMHTAVSAVHAZ 122
QY 126 DGRITHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDQNGYRF 185
DB 123 NGEALLAGAVIDGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDQNGYRF 182
QY 186 VYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGNPLOTLLREEGAL 245
DB 183 VYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGNPLOTLLREEGAL 242
QY 246 GDNQWQOQGFPMNLNMLFLAGPAESRWVQRYGFLPDLIARFYAGKLTVDRLR 305
DB 243 VTSRLCGPMRREASGMRAGLFHPTTGYSLPLAVADRLSALDVFTSSVHCTI 302
QY 306 ORWQOQGFPMNLNMLFLAGPAESRWVQRYGFLPDLIARFYAGKLTVDRLR 365
DB 303 RHWKQGFPMRREASGMRAGLFHPTTGYSLPLAVADRLSALDVFTSSVHCTI 362

RESULT 3
US-08-096-623A-14
; Sequence 14, Application US/08096623A
; Patent No. 5684238
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Huel-Che B.
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
; TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welsh & Katz, Ltd.
; STREET: 120 S. Riverside Plaza, 22nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,623A
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,061
; FILING DATE: 09-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,921

FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA: US 07/562,674
APPLICATION NUMBER: 03-AUG-1990
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-096-623A-14

Query Match 55.8%; Score 1127; DB 1; Length 374;
Best Local Similarity 58.4%; Pred. No. 1.7e-104;
Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

6 DLILVAGLANGLIALRLOOHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIAPL 65
3 DLILVAGLANGLIALRLOOHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIAPL 62
66 VVHEWPDYQVRRPQRRHNSGYCVTSRHFAGILRQOFGHLLWHTAVSAVHAESVOLA 125
63 VAHWAPGYEQVDPDLRLRARGYSITSEFALHQAALGENIWNCSVSEVLPSVRLA 122
126 DGRITIIHASTVIDGRGTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQNGYRF 185
123 NGEALLAGAVIDGRGTVTASSAMOTGYQLFGQWRLTQPHGLTVPILMDATVAQOQGYRF 182
186 VYTLPLSATALLIEDTHYIDKANLQAEARQNIQDYAARQGMPLQTLLEEGCALPITLT 245
183 VYTLPLSATALLIEDTHYIDKANLQAEARQNIQDYAARQGMPLQTLLEEGCALPITLT 242
246 GDNRFQWQOQOQAGSLRAGLEFHTPTGYSIPLAVALADRLSALDVFTSSVHQTIAHFAQ 305
243 VTSRLCGPMRRAASGNRAGLFHTPTGYSIPLAVALADRLSALDVFTSSVHQTIAHFAQ 302
306 QRWQOQGFEMKLNMLFLAGPAESRWVMOQFVGLPDLIARYAGKLTVDRLRLSOK 365
303 RHWPRQGFPELLNMLFLAGPAESRWVMOQFVGLPDLIARYAGKLTVDRLRLSOK 362
366 PPVPV 370
363 PPVPL 367

RESULT 4
IS-08-660-645A-7
Sequence 7, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: PERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street

CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-660-645A-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

5 YDLILVAGLANGLIALRLOOHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIAPL 63
3 HDLLIAGAGUSGALLAVALADRLSALDVFTSSVHQTIAHFAQ 59
64 --PLVWHPDYQVRRPQRRHNSGYCVTSRHFAGILRQOFGHLLWHTAVSAVHAES 121
60 LSPRRGEWTDQVAFPHDSRLTGTGYSIEAGALIGLLQ---GVDLRWNTHTVATLDDTG 116
122 VOLADGRITIIHASTVIDGRGTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQON 181
117 ATLDGSRIRIACVAVIDARGAVETPHLTGQKFGVVEIETDAPHGVERPMIMDATVPMQD 176
182 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIQDYAARQGMPLQTLLEEGCALP 241
177 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIQDYAARQGMPLQTLLEEGCALP 235
242 ITLGDNRQFQWQOQOQAGSLRAGLEFHTPTGYSIPLAVALADRLSALDVFTSSVHQT 299
236 IALAADAIGFWRDHAQGVAVPVLGAGLFHTPTGYSIPLAVALADRLSALDVFTSSVHQT 294
300 IAHFAQOQWQOQ 359
295 VGVWADRADRDRLNMLFLAGPAESRWVMOQFVGLPDLIARYAGKLTVDRL 354
360 RILSGKPPVVPVFAALQAI 377
355 RIVTGRPPPLSQAVRCL 372

RESULT 5
US-09-298-718-7
Sequence 7, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-298-718-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
QY 5 YDLILVGLAGLIALRLQOQHDPMDRIILLIEAGPEAGGNHTSFHEDLTINQHRWIA- 63
DB 3 HDLLIAGAGLSGALIALAVRRDPDARIWMLDARSQSDQHTWSCHTDLS---PEWLAR 59
QY 64 --PLVVHWPDYQVRRPQRRRHVNSGYCVTSRHFAGILRQOFGHILWLTAVSAVHAES 121
DB 60 LSPFIRRGWTDQEVAFPDHSRRLTTGYGSIIEAGALIGLLQ---GVDLRWNTHVATLDDTG 116
QY 122 VOLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPGLSSPIIMDATVDDQN 181
DB 117 ATLTDGSRIEACVIDARGAVETPHLTGVFKFVGVEIETDAPHGVERPMTDATVPQMD 176
QY 182 GYRFVYTLPLSATALLIEDTHYIDKANIQARERQNIIDYAAARQGWPLQTLLEEQGALP 241
DB 236 IALAHDAICFWRDHAQAGVAVPVGLAGLPHFVTVGSLPVAQAQVADAIARDL-TTASARRA 294
QY 300 IAHFAQQRQWQOGPFRLNRLMFLAGPAESRWYRMYRFGYGLPEDIARFYAGKLTVDRL 359
DB 295 VRGNAIDRADRDRLNRLMFLRCCPPDRYRLRQFYRLPQPLIERFYAGRLTLADRL 354
QY 360 RILSGKPPVPVFAALQAI 377
DB 355 RIVTGRPPILSQVRCL 372

RESULT 6

US-09-546-969-7

Sequence 7, Application US/09546969
Patent No. 6207409
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546,969
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-546-969-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
QY 5 YDLILVGLAGLIALRLQOQHDPMDRIILLIEAGPEAGGNHTSFHEDLTINQHRWIA- 63
DB 3 HDLLIAGAGLSGALIALAVRRDPDARIWMLDARSQSDQHTWSCHTDLS---PEWLAR 59
QY 64 --PLVVHWPDYQVRRPQRRRHVNSGYCVTSRHFAGILRQOFGHILWLTAVSAVHAES 121
DB 60 LSPFIRRGWTDQEVAFPDHSRRLTTGYGSIIEAGALIGLLQ---GVDLRWNTHVATLDDTG 116
QY 122 VOLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPGLSSPIIMDATVDDQN 181
DB 117 ATLTDGSRIEACVIDARGAVETPHLTGVFKFVGVEIETDAPHGVERPMTDATVPQMD 176
QY 182 GYRFVYTLPLSATALLIEDTHYIDKANIQARERQNIIDYAAARQGWPLQTLLEEQGALP 241
DB 177 GYRFYLLPFPSTRIILIEDTRYSDGDLDDCALAQASLDYAAARGWTGQE-MRREGILP 235
QY 242 ITLTGDNRQFWQQOQAC--SGLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSSVHQT 299
DB 236 IALAHDAICFWRDHAQAGVAVPVGLAGLPHFVTVGSLPVAQAQVADAIARDL-TTASARRA 294
QY 300 IAHFAQQRQWQOGPFRLNRLMFLAGPAESRWYRMYRFGYGLPEDIARFYAGKLTVDRL 359
DB 295 VRGNAIDRADRDRLNRLMFLRCCPPDRYRLRQFYRLPQPLIERFYAGRLTLADRL 354
QY 360 RILSGKPPVPVFAALQAI 377

355 RIVTGRPPPLSQAVRCL 372

RESULT 7

S-08-980-832-5
Sequence 5, Application US/08980832B
Patent No. 6291204
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/08/980,832B
CURRENT FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 5
LENGTH: 382
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
S-08-980-832-5

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
Y 5 YDILIVGAGLANGLIARLQQQHPDNRILLIIBAGPEAGNHTWSFHEBDLTINQHRWIA- 63
b 3 HDLLIAGAGLSGALIALAVDRRDPARIWMLDARSQPSDQHTWSCHDTLS---PEWLAR 59
Y 64 --FLVVHHPDYVRFPQRRRHVNSGYCVTSRHFAGILRQQFGQHLMLHTAVSAVHAES 121
b 60 LSFIRGEWTDQEVAFPDHSRRLTTGYSIEAGALIGLLQ---GVDLRNTHVATLDDTG 116
Y 122 VQADGRIIHASTVIDRGYTPDSALRVGFQAFI GOEWOLSAFGLSSPIIMDATVDQON 181
b 117 ATLTGDSRIEACVIDARGAVETPHLTVGQFVGVEIETDAPHGVERPMIMDATVPQMD 176
Y 182 GYEFVYTLPLSATALLIETHYIDKANLQAEARQNI RDYAARQGWPLQTLREEGALP 241
b 177 GYRPIYLLPSPRIILIEDTRYSDGGLDDGALAAQSLDYAARRGTGQS-MRRERGILP 235
Y 242 ITLTGDNROFWOQPOAC--SGLRAGLPFTTGYSLPLAVALADRLSALDVTSSVHQT 299
b 236 IALAHDAIGFWRDHAQGVAVPGLGAGLFPHVTGYSILPYAAQVADAIAARDL-TTASARRA 294
Y 300 IAHFAQORWQQQGFPMNMLFLAGPAESRVRVQRFYGLPBDLIIARFYAGKLTVTDR 359
b 295 VRGWAIDRADRDRLFLNMLFRGCPDPRYRLQLRFRYLPQPLIERFYAGRLTLADRL 354
Y 360 RILSGKPPVPVFAALQAI 377
b 355 RIVTGRPPPLSQAVRCL 372

RESULT 8

S-09-547-267-7
Sequence 7, Application US/09547267
Patent No. 6613543
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA

ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-547-267-7

Query Match 38.8%; Score 783.5; DB 4; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
Qy 5 YDILIVGAGLANGLIARLQQQHPDNRILLIIBAGPEAGNHTWSFHEBDLTINQHRWIA- 63
Db 3 HDLLIAGAGLSGALIALAVDRRDPARIWMLDARSQPSDQHTWSCHDTLS---PEWLAR 59
Qy 64 --FLVVHHPDYVRFPQRRRHVNSGYCVTSRHFAGILRQQFGQHLMLHTAVSAVHAES 121
Db 60 LSFIRGEWTDQEVAFPDHSRRLTTGYSIEAGALIGLLQ---GVDLRNTHVATLDDTG 116
Qy 122 VQADGRIIHASTVIDRGYTPDSALRVGFQAFI GOEWOLSAFGLSSPIIMDATVDQON 181
Db 117 ATLTGDSRIEACVIDARGAVETPHLTVGQFVGVEIETDAPHGVERPMIMDATVPQMD 176
Qy 182 GYEFVYTLPLSATALLIETHYIDKANLQAEARQNI RDYAARQGWPLQTLREEGALP 241
Db 177 GYRPIYLLPSPRIILIEDTRYSDGGLDDGALAAQSLDYAARRGTGQS-MRRERGILP 235
Qy 242 ITLTGDNROFWOQPOAC--SGLRAGLPFTTGYSLPLAVALADRLSALDVTSSVHQT 299
Db 236 IALAHDAIGFWRDHAQGVAVPGLGAGLFPHVTGYSILPYAAQVADAIAARDL-TTASARRA 294
Qy 300 IAHFAQORWQQQGFPMNMLFLAGPAESRVRVQRFYGLPBDLIIARFYAGKLTVTDR 359
Db 295 VRGWAIDRADRDRLFLNMLFRGCPDPRYRLQLRFRYLPQPLIERFYAGRLTLADRL 354
Qy 360 RILSGKPPVPVFAALQAI 377
Db 355 RIVTGRPPPLSQAVRCL 372

RESULT 9

US-09-920-923B-5
Sequence 5, Application US/09920923B
Patent No. 6677134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920,923B
CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in version 3.1
SEQ ID NO. 5
LENGTH: 382
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923B-5

Query Match 38.8%; Score 783.5; DB 4; Length 382;
Best Local Similarity 43.7%; Score No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
QY 5 YDLILVAGLANGLIALRLQOQHDPMDRIILLIEAGPEAGNHTWSFHEEDLTNLQHRWI 63
DB 3 HDLLIAGAGLIALAVDRDPDARIWMLDARSQSDQHTWSCHTDLS--PEWLAR 59
QY 64 --PLVVHHPDYQVRPQRRHNSGYCVTSRHPAGILRQOFGQHLHLHTAVSAVHA 121
DB 60 LSPFRGEWTDEVAFPDHSRLTTGYSGIERGALIGLQ--GVDLWNTHVATLDDTG 116
QY 122 VOLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHSLSPIIMDATVQDN 181
DB 117 ATLTDGSRTEAACVIDARGAVETPHLTGVGFQVGEIETDAPHGVERPMIMDATVQMD 176
QY 182 GYRFVYTLPLSATALIETHYIDKANIQAERARQNIQDYAARQWPLQTLRLREOQALP 241
DB 177 GYRFVYTLPLSATALIETHYIDKANIQAERARQNIQDYAARQWPLQTLRLREOQALP 241
QY 242 IYLTGDNRFQWQOQAC--SGRLAGLFHTTGYSLPLAVALADRLSALDVTSSSVHQT 299
DB 236 IALAHDAIGFWRDHAQAVPVGLGAGLFHTTGYSLPLAVALADRLSALDVTSSSVHQT 294
QY 300 IAHFAQORWQOQGFPMNLNMLFLAGPAESRWVQRYGCPEDLIARFYAGKLTVTDL 359
DB 295 VRGWADRDRDFRLNRLMFRGCPDRYLLQRFYLPQFLRFRYAGRLTLADRL 354
QY 360 RILSGKPPVPVPAALQAI 377
DB 355 RIVTGRPIPLGTAIRCL 372

RESULT 10
US-08-663-310-6
Sequence 6, Application US/08663310
Patent No. 5811273
GENERAL INFORMATION:
APPLICANT: MISAHA, No. 5811273ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
TITLE OF INVENTION: XANTHOPHYLLS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,310
FILING DATE: 23-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-663-310-6
Query Match 37.6%; Score 759.5; DB 2; Length 386;
Best Local Similarity 43.0%; Pred No. 1.1e-67;
Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;
QY 5 YDLILVAGLANGLIALRLQOQHDPMDRIILLIE--AGPEAGNHTWSFHEEDLTNLQHRWI 62
DB 3 HDVLLAGAGLIALALRAARPOLRVLILLDHAAGPSDG--HTWSCHDPLSDP--WL 57
QY 63 A--PLVVHHPDYQVRPQRRHNSGYCVTSRHPAGILRQOFGQHLHLHTAVSAVHA 119
DB 58 ALKPLRANWPDQVFRPHARRLATYGLSDGAALADAVRSGAEIRW--DSDIALLDA 116
QY 120 ESQLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHSLSPIIMDATVQDN 179
DB 117 OGATLSCGTRIEAGAVLDGCGAQPSEHLTVGFQVGEIETDPRPHGVPREIMDATVQ 176
QY 180 ONGYRFVYTLPLSATALIETHYIDKANIQAERARQNIQDYAARQWPLQTLRLREOQAL 239
DB 177 QGYRFVYTLPLSATALIETHYIDKANIQAERARQNIQDYAARQWPLQTLRLREOQAL 235
QY 240 LPIITLTDGDNRFQWQOQAC--QACSLRAGLFHTTGYSLPLAVALADRLSALD--VFTSSSV 296
DB 236 LPIALAHDAIGFWRDHAQAVPVGLGAGLFHTTGYSLPLAVALADRLSALD--VFTSSSV 295
QY 297 HOTIAHFAQORWQOQGFPMNLNMLFLAGPAESRWVQRYGCPEDLIARFYAGKLTVT 356
DB 296 RGAIRDYADRDRDFRLNRLMFRGCPDRYLLQRFYLPQFLRFRYAGRLTLADRL 355
QY 357 DRLRILSGKPPVPVPAALQAI 377
DB 356 DQLRIVTGRPIPLGTAIRCL 376

RESULT 11
US-09-006-491-6
Sequence 6, Application US/09006491
Patent No. 5972690
GENERAL INFORMATION:
APPLICANT: MISAHA, No. 5972690ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
TITLE OF INVENTION: XANTHOPHYLLS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/006,491

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

TELEPHONE:

TELEFAX:

RESULT 12

US-09-335-919-6
; Sequence 6, Application US/09335919
; Patent No. 6150130
; GENERAL INFORMATION:
; APPLICANT: MISA, No. 6150130:hiko
; APPLICANT: KONDO, Keiji
; APPLICANT: KAJIMARA, Susumu
; APPLICANT: YOKOYAMA, Akihiro
; TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
; TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
; NUMBER OF INVENTIONS: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/335,919
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US/08/663,310
; FILING DATE: 23-SEP-1996
; APPLICATION NUMBER: WO PCT/JP94/02220
; FILING DATE: 26-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-235917
; FILING DATE: 05-SEP-1994
; PRIOR APPLICATION DATA: JP 5-348737
; FILING DATE: 27-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 49441/109
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-335-919-6

Query Match 37.6%; Score 759.5; DB 3; Length 386;

Best Local Similarity 43.0%; Pred. No. 1.1e-67;

Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;

Qy 5 YDLILVAGLANGLIALRLQQQHPDMRIILLIE--AGPEAGNHTWSFHEEDLTINQHRWI 62
Db 3 HDVLLAGLANGLIALALRAARPDRLVLLDHAAGPSDG--HTWSCHDPDLSPD---WL 57
Qy 63 A---PLVHHWPDYQVRFPORRHVNSGYCYTSRHFAGILRQFGQHLWLHTAVSAVHA 119
Db 58 ARLKPLRRANWPDQVRFPPHARRLATGYSLDGAALADAVVRSAGAEIRW--DSDIALDLA 116
Qy 120 ESQLADGRIIHAISTVIDRGVTPDSALRVGFQAFIQTGQWQLSAPHGLSSPTIMDATVDQ 179
Db 117 QGATLSCGTRIIEAGVLDGRGAQPSRHLTVGFQKFGVGEIETDRPHGVPRMIMDATVTQ 176
Qy 180 QNGYRVYTLPLSATALLIEDTHYIDKANLQERARQNRIDYARQGWPLQTLREBOGA 239

S-09-006-491-6

Query Match 37.6%; Score 759.5; DB 2; Length 386;

Best Local Similarity 43.0%; Pred. No. 1.1e-67;

Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;

Y 5 YDLILVAGLANGLIALRLQQQHPDMRIILLIE--AGPEAGNHTWSFHEEDLTINQHRWI 62
b 3 HDVLLAGLANGLIALALRAARPDRLVLLDHAAGPSDG--HTWSCHDPDLSPD---WL 57
Y 63 A---PLVHHWPDYQVRFPORRHVNSGYCYTSRHFAGILRQFGQHLWLHTAVSAVHA 119
b 58 ARLKPLRRANWPDQVRFPPHARRLATGYSLDGAALADAVVRSAGAEIRW--DSDIALDLA 116
Y 120 ESQLADGRIIHAISTVIDRGVTPDSALRVGFQAFIQTGQWQLSAPHGLSSPTIMDATVDQ 179
b 117 QGATLSCGTRIIEAGVLDGRGAQPSRHLTVGFQKFGVGEIETDRPHGVPRMIMDATVTQ 176
Y 180 QNGYRVYTLPLSATALLIEDTHYIDKANLQERARQNRIDYARQGWPLQTLREBOGA 239
b 177 QNGYRVYTLPLSATALLIEDTHYIDKANLQERARQNRIDYARQGWPLQTLREBOGA 239
Y 240 LPITLTDNRQFQQOOP--QACSGELAGLHFTPTGYSLPLAVALADRLSALD-VFTSSV 296
b 236 LPITALHDAAGFADHAAQFVPLVGLRAGFHEVPTGYSLPYAAQVADVAGLSGPGTDAL 295
Y 297 HQTIAHQAQRWQQOQFFFWLNRMLFLAGPAESNRVWQRFVGLPEDLIAPYAGKLTVT 356
b 296 RGAIRDYADRAKRDREKLLNRMLFRGCAPDRRTLLQRFYRMPHGLIBRFYAGRLSVA 355
Y 357 DRLRILSGKPPVPVFAALQAI 377
b 356 DQLRIVTGKPPPLGTAIRCL 376

Db 177 ODGYRFLYLLPSPTRILIEDTRYSDGGDLDDALAAASHDYARQQOW--TCAEVRREGEI 235
QY 240 LPITUTGNORFQWQOP--QACSLRAGLPHPTTGYSLPLAVALADRLSALD-VFTSSSV 296
Db 236 LPIALAHDAAGFADHAGPVPVGLRAGFFPVTGYSLPAAQVADVAGLSGPPGTDAL 295
QY 297 HOTIAHFAQQWQOGFFRMLNMLFLAGPAESRWVWQRFYGPULPEDIARFVAGKLTUT 356
Db 296 RGAIKDYAIDARRDRFLRLNMLPRGCDPRRYTLQRFYRPHGLIERFVAGRLSVA 355
QY 357 DELRILSGKPPVPVFAALQAI 377
Db 356 DQLRIVTGKPPPIUGTAIRCL 376

RESULT 13

US-09-134-607A-19
Sequence 19, Application US/09134607A

Patent No. 6252141

GENERAL INFORMATION:

APPLICANT: Joseph Hirschberg et al.

TITLE OF INVENTION: POLYNUCLEOTIDES CONTROLLING THE EXPRESSION

TITLE OF INVENTION: OF AND CODING FOR GENE B IN TOMATO AND USE

TITLE OF INVENTION: OF SAME FOR ALTERING CAROTENOID

TITLE OF INVENTION: BIOSYNTHESIS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

STREET: 20001 Jefferson Davis Highway, Suite 207

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead, Slimnote 890TX

OPERATING SYSTEM: MS DOS version 6.2,

OPERATING SYSTEM: Windows version 3.11

SOFTWARE: Word for Windows version 2.0,

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134,607A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 325/12

TELECOMMUNICATION INFORMATION:

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX:

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 498

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-134-607A-19

Query Match 7.6%; Score 154.5; DB 3; Length 498;

Best Local Similarity 22.1%; Pred. No. 6.3e-07;

Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

QY 5 YDLILVAGLANGLIALRLOQO--HPDMRILLIEAGPEA---GGNHTWSPFHEEDLTN-- 57

Db 82 FDVILIGAGPA---GLRLAEQSVKYGKVCVDPSPLSMFWNNYGVWVDFENIGLEDC 137

QY 58 -QHRVAPLVWHPDYQVRP---PQRR-----RHVNSGYCVTSRHPAGILRQOFG 105

Db 82 FDVILIGAGPA---GLRLAEQSVKYGKVCVDPSPLSMFWNNYGVWVDFENIGLENC 137

Db 138 LDHKV--PMTCVHINDNKTGILGRPYGRVSEKKLKLKLLNS---CVENR-----VKFYK 186
QY 106 QHLWLTAVSAVHAE---SVQLADGRIIHASTVIDGRGYTP-----DSALRVGFAQFIG 157
Db 187 AKVW-----KVEHEPFESSIVCDDGKKIRGSLWVDSGFSADFIEYDKPRNHGYQIAHGV 241
QY 158 ENQL--SAPHGSLSSPLIND-----ATVDCQNGYRFVYTLPLSATALLIEDTHYID 205
Db 242 LVEVNHHPFDLKKVMDWRDRLSHGNPEYLVNNAKEPTFLYAMPFDRNLVFLBETSLVS 301
QY 206 KANLQAEARQNIQDYAARQGMPLQTLRLREQGALPITLTGDNRFQWQOQFQACSGLRAG 265
Db 302 RPLVSYMEVXRMRVARLRHLGKIVRSVIEEKVCVPM---GGPLRIPQNVMAIGG-NSG 357
QY 266 LPHPTTGYSLPLAVALADRLS 286
Db 358 IVPSTGYMVARSMALAPVLA 378

RESULT 14

US-09-134-607A-17

Sequence 17, Application US/09134607A

Patent No. 6252141

GENERAL INFORMATION:

APPLICANT: Joseph Hirschberg et al.

TITLE OF INVENTION: POLYNUCLEOTIDES CONTROLLING THE EXPRESSION

TITLE OF INVENTION: OF AND CODING FOR GENE B IN TOMATO AND USE

TITLE OF INVENTION: OF SAME FOR ALTERING CAROTENOID

TITLE OF INVENTION: BIOSYNTHESIS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

STREET: 20001 Jefferson Davis Highway, Suite 207

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead, Slimnote 890TX

OPERATING SYSTEM: MS DOS version 6.2,

OPERATING SYSTEM: Windows version 3.11

SOFTWARE: Word for Windows version 2.0,

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134,607A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 325/12

TELECOMMUNICATION INFORMATION:

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX:

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 498

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-134-607A-17

Query Match 7.5%; Score 152.5; DB 3; Length 498;

Best Local Similarity 22.1%; Pred. No. 1e-06;

Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

QY 5 YDLILVAGLANGLIALRLOQO--HPDMRILLIEAGPEA---GGNHTWSPFHEEDLTN-- 57

Db 82 FDVILIGAGPA---GLRLAEQSVKYGKVCVDPSPLSMFWNNYGVWVDFENIGLENC 137

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:26:38 ; Search time 45.165 Seconds
(without alignments)
2389.754 Million cell updates/sec

title: US-09-941-947A-30

effect score: 2021

sequence: 1 MQPHYDLILVAGLGLA.....SGKPPVPVFALQAMTTHR 382

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

A Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2021	100.0	382	5	Aae22313 Pantoea s
2	2021	100.0	382	6	Aao16019 Pantoea s
3	2021	100.0	382	6	Abp96687 Pantoea s
4	1678	83.0	382	2	Aar07465 Polypepti
5	1678	83.0	382	2	Aaw82258 C. utilis
6	1678	83.0	382	2	Aaw87891 Protein e
7	1678	83.0	382	2	Aaw99100 Erwinia u
8	1127	55.8	374	2	Aaw01125 Lycopen
9	1127	55.8	374	2	Aaw32474 Erwinia h
10	1126	55.7	374	2	Aar13987 Lycopen
11	1040.5	51.5	392	6	Abm70122 Phototrab
12	783.5	38.8	382	2	Aaw06517 Flavobact
13	783.5	38.8	382	2	Aaw69533 Flavobact
14	759.5	37.6	386	2	Aar79062 3 hydroxy
15	759.5	37.6	386	2	Aaw87884 Protein e
16	759.5	37.6	386	2	Aaw99096 Agrobacte
17	637	31.5	434	2	Aar95698 Erythroba
18	154.5	7.6	376	6	Aae31691 Rhodococc
19	154.5	7.6	498	3	Aay70397 Protein e
20	153.5	7.6	411	2	Aar76980 Lycopen
21	152.5	7.5	498	3	Aay70395 Protein e
22	152.5	7.5	498	3	Aay70395 Protein e
23	152	7.5	526	3	Aay54309 Amino aci
24	148.5	7.3	411	2	Aaw00276 Cyanobact
25	148.5	7.3	498	2	Aar82242 Capsanthi

26	148.5	7.3	501	3	AAY54311	Amino aci
27	148.5	7.3	501	3	AAG53411	Arabidops
28	148.5	7.3	501	5	ABBS2263	Herbicida
29	142.5	7.1	502	3	AAY54312	Amino aci
30	141.5	7.0	503	3	AAY54317	Amino aci
31	139.5	6.9	369	3	AAG53412	Arabidops
32	139.5	6.9	511	3	AAY90226	Marigold
33	139.5	6.9	511	3	AAY54316	Amino aci
34	139	6.9	498	2	AAW06454	Capsicum
35	139	6.9	498	3	AAY54313	Amino aci
36	139	6.9	500	3	AAY54315	Amino aci
37	136.5	6.8	517	4	AAE09798	Spinach 1
38	136	6.7	328	3	AAG53413	Arabidops
39	134.5	6.7	516	3	AAY90228	Marigold
40	134.5	6.7	516	3	AAY54310	Amino aci
41	133	6.6	407	6	ABU39747	Protein e
42	133	6.6	500	3	AAY54314	Amino aci
43	132.5	6.6	456	3	AAY54289	Consensus
44	131.5	6.5	524	5	ABG32343	Chimaeric
45	130.5	6.5	529	3	AAY54307	Amino aci

ALIGNMENTS

RESULT 1

AAE22313

ID AAE22313 standard; protein; 382 AA.

AC AAE22313;

XX

DT 25-JUL-2002 (first entry)

DE Pantoea stewartii lycopene cyclase (CrtY) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; lycopene cyclase; CrtY.

XX Pantoea stewartii.

OS

XX WO200218617-A2.

PN

XX 07-MAR-2002.

PD

XX 04-SEP-2001; 2001WO-US027420.

PF

XX 01-SEP-2000; 2000US-0229858P.

PR

XX 01-SEP-2000; 2000US-0229907P.

PR

XX (DUPO) DU PONT DE NEMOURS & CO E I.

PA

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odom JW, Picataggio SK, Rouviere PE;

PI WPI; 2002-351711/38.

XX N-PSDB; AAD35511.

DR

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates.

XX

PS Claim 21; Page 139-140; 156pp; English.

XX

XX The invention relates to a method for producing carotenoid compounds. The

XX method comprises a transformed metabolising host cell, comprising

XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX control of regulatory sequences, and contacting the host cell with carbon

XX substrate to produce a carotenoid compound. The method is useful for

XX producing carotenoid compounds such as antheraxanthin and astaxanthin, by

XX using microorganism having a nucleic acid molecule encoding enzymes in

XX

CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Pantoea stewartii lycopene cyclase (CrtY) enzyme used in the invention
 XX
 SQ Sequence 382 AA;

Query Match 100.0%; Score 2021; DB 5; Length 382;
 Best Local Similarity 100.0%; Pred. No. 8.1e-192;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 DB 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 QY 61 WTAPLVVHHWPDYQVRFPPQRRRHVNSGYCVTSRHFAGILRQOFGQHLWHTAVSAVHAE 120
 DB 61 WTAPLVVHHWPDYQVRFPPQRRRHVNSGYCVTSRHFAGILRQOFGQHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHAFTVDRGYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDQO 180
 DB 121 SVQLADGRIIHAFTVDRGYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDQO 180
 QY 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 DB 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 QY 241 PITLGDNRQFWQOQPOACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVTSSVHQTI 300
 DB 241 PITLGDNRQFWQOQPOACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVTSSVHQTI 300
 QY 301 AHFAQORWQOQGFRRMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 DB 301 AHFAQORWQOQGFRRMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 2
 AAO16019
 ID AAO16019 standard; protein; 382 AA.
 XX
 AC AAO16019;
 XX
 DT 20-FEB-2003 (first entry)
 XX
 DE Pantoea stewartii lycopene Beta-cyclase.
 XX
 XX Carotenoid; crt.
 XX
 XX Pantoea stewartii.
 XX
 PN WO200273995-A2.
 XX
 XX 10-OCT-2002.
 XX
 XX 25-JAN-2002; 2002WO-US002124.
 XX
 XX 26-JAN-2001; 2001US-0264329P.
 XX
 XX 04-MAY-2001; 2001US-028984P.
 XX
 XX (CRGI) CARGILL INC.
 XX
 XX De Souza ML, Kollmann SR, May CA, Schroeder WA;
 XX
 XX WPI; 2003-075455/07.
 XX
 XX N-PSDB; ABT14191.
 XX
 XX Novel isolated nucleic acid useful e.g. to engineer host cells with the

PT ability to produce particular carotenoids and polypeptides useful in cell
 PT -free systems to make particular carotenoids.

PS Claim 14; Page 60-61; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
 CC carotenoid (crt)-related proteins. The crt-related DNA and protein
 CC sequences of the invention are useful for engineering cells which are
 CC able to produce carotenoids. The present amino acid sequence represents a
 CC crt-related protein of the invention

XX SQ Sequence 382 AA;

Query Match 100.0%; Score 2021; DB 6; Length 382;
 Best Local Similarity 100.0%; Pred. No. 8.1e-192;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 DB 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 QY 61 WTAPLVVHHWPDYQVRFPPQRRRHVNSGYCVTSRHFAGILRQOFGQHLWHTAVSAVHAE 120
 DB 61 WTAPLVVHHWPDYQVRFPPQRRRHVNSGYCVTSRHFAGILRQOFGQHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHAFTVDRGYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDQO 180
 DB 121 SVQLADGRIIHAFTVDRGYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDQO 180
 QY 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 DB 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 QY 241 PITLGDNRQFWQOQPOACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVTSSVHQTI 300
 DB 241 PITLGDNRQFWQOQPOACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVTSSVHQTI 300
 QY 301 AHFAQORWQOQGFRRMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 DB 301 AHFAQORWQOQGFRRMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 3
 ABP96687
 ID ABP96687 standard; protein; 382 AA.
 XX
 AC ABP96687;
 XX
 DT 03-JUN-2003 (first entry)
 XX
 DE Pantoea stewartii lycopene cyclase SEQ ID NO:6.
 XX
 XX Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
 KW crtI; crtB; crtZ; lycopene cyclase; enzyme; phytoene; carotenoid.
 XX
 XX Pantoea stewartii.
 OS
 XX WO2003016503-A2.
 XX
 XX 27-FEB-2003.
 XX
 XX 15-AUG-2002; 2002WO-US026647.
 XX
 XX 15-AUG-2001; 2001US-0312646P.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;
 XX

R WPI: 2003-268323/26.
 X N-PSDB; ACC44761.
 X Novel nucleic acid molecule isolated from *Pantoea stewartii* encoding a
 T carotenoid biosynthetic enzyme, useful for regulating carotenoid
 T biosynthesis in an organism.
 X Claim 4; Page 61-62; 68pp; English.
 S
 C The present invention describes *Pantoea stewartii* carotenoid biosynthetic
 C enzymes (I). More specifically described are the geranylgeranyl
 C pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
 C lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
 C (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see AB996685 to
 C AB996690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
 C carotenoid biosynthesis in an organism, by over-expressing (I) in an
 C organism, such that the carotenoid biosynthesis is altered in the
 C organism. (I) and the genes encoding (I) are useful for converting
 C phytoene to the carotenoids, for creating recombinant organisms that have
 C the ability to produce various carotenoid compounds, and also for
 C enhancing or manipulating carotenoid compounds. (I) can also be used for
 C producing gene products having enhanced or altered activity
 X
 X Sequence 382 AA;

Query Match 100.0%; Score 2021; DB 6; Length 382;
 Best Local Similarity 100.0%; Pred. No. 8.1e-192;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1 MOPHYDLILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHR 60
 b 1 MOPHYDLILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHR 60
 Y 61 WIAPLVVHEWPDYQVFPQRRHVNNGYCVTSRHPAGILRQOFGCHLWLTAVSAVHAE 120
 b 61 WIAPLVVHEWPDYQVFPQRRHVNNGYCVTSRHPAGILRQOFGCHLWLTAVSAVHAE 120
 Y 121 SVOLADGRIIHASTVIDGRGTPDSALRVGFQAFIQEWOLSAFGLSSPIIMDATVDQO 180
 b 121 SVOLADGRIIHASTVIDGRGTPDSALRVGFQAFIQEWOLSAFGLSSPIIMDATVDQO 180
 Y 181 NGRVFVYVTLPLSATALLEDTHYIDKANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
 b 181 NGRVFVYVTLPLSATALLEDTHYIDKANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
 Y 241 PITLTGDNROFWOQOQACSLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSVHQTI 300
 b 241 PITLTGDNROFWOQOQACSLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSVHQTI 300
 Y 301 AHFAQORWQOQGFRRMLNRLMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
 b 301 AHFAQORWQOQGFRRMLNRLMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
 Y 361 ILSGKPPVPVFAALQAIMTTHR 382
 b 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 4
 AR07465
 D - AAR07465 standard; protein; 382 AA.
 X C AAR07465;
 X T 24-OCT-2003 (revised)
 T 28-JAN-1991 (first entry)
 X Polypeptide with enzymatic activity for the conversion of lycopene into
 E beta-carotene.
 X Carotenoid biosynthesis; vitamin A; cancer; food coloring.
 X *Pantoea ananatis*.

XX BP393690-A.
 PN 24-OCT-1990.
 XX 20-APR-1990; 90BP-00107493.
 XX 21-APR-1989; 89JP-00103078.
 PR 05-MAR-1990; 90JP-00053255.
 XX (KIRI) KIRIN BEER KK.
 PA Misawa N, Kobayashi K, Nakamura K;
 XX WPI; 1990-322212/43.
 DR N-PSDB; AAQ06295.
 XX DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn.
 PT of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.
 XX Claim 3; Fig 3; 40pp; English.
 XX Gene products are useful for the synthesis of carotenoids, useful as food
 CC coloring, vitamin A precursor, and possibly in prevention of cancer. See
 CC also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)
 XX Sequence 382 AA;

Query Match 83.0%; Score 1678; DB 2; Length 382;
 Best Local Similarity 82.2%; Pred. No. 1e-157;
 Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;
 Qy 1 MOPHYDLILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHR 60
 Db 1 MOPHYDLILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHR 60
 Qy 61 WIAPLVVHEWPDYQVFPQRRHVNNGYCVTSRHPAGILRQOFGCHLWLTAVSAVHAE 120
 Db 61 WIAPLVVHEWPDYQVFPQRRHVNNGYCVTSRHPAGILRQOFGCHLWLTAVSAVHAE 120
 Qy 121 SVOLADGRIIHASTVIDGRGTPDSALRVGFQAFIQEWOLSAFGLSSPIIMDATVDQO 180
 Db 121 SVOLADGRIIHASTVIDGRGTPDSALRVGFQAFIQEWOLSAFGLSSPIIMDATVDQO 180
 Qy 181 NGRVFVYVTLPLSATALLEDTHYIDKANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
 Db 181 NGRVFVYVTLPLSATALLEDTHYIDKANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
 Qy 241 PITLTGDNROFWOQOQACSLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSVHQTI 300
 Db 241 PITLTGDNROFWOQOQACSLRAGLFHPTTGYSLPLAVLADRLSALDVFTSSVHQTI 300
 Qy 301 AHFAQORWQOQGFRRMLNRLMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
 Db 301 AHFAQORWQOQGFRRMLNRLMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
 Qy 361 ILSGKPPVPVFAALQAIMTTHR 382
 Db 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 5
 AAW82258
 ID AAW82258 standard; protein; 382 AA.
 XX AAW82258;
 AC AAW82258;
 XX 17-OCT-2003 (revised)
 DT 16-JUL-1999 (first entry)
 XX C. utilis crtY protein.
 XX HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtY;

KW carotenoid.
 XX Pichia jadinii.
 XX JP10248575-A.
 XX 22-SEP-1998.
 XX 12-MAR-1997; 97JP-00058012.
 XX 12-MAR-1997; 97JP-00058012.
 XX (KIRI) KIRIN BREWERY KK.
 XX WPI; 1998-560727/48.
 XX N-PSDB; AAV73182.
 XX Gene useful for increase in carotenoid production - and preparation of carotenoid.
 XX Example 2; Fig 15-17; 54pp; Japanese.
 XX This invention describes a novel method for the preparation of carotenoids using genes and proteins isolated from Candida utilis. The invention specifically describes the isolation of a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence represents the Candida utilis crtY protein which is used in the method of the invention. (Updated on 17-OCT-2003 to standardise OS field)
 XX SQ Sequence 382 AA;
 Query Match 83.0%; Score 1678; DB 2; Length 382;
 Best Local Similarity 82.2%; Pred. No. 1e-157;
 Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;
 QY 1 MOPHYDLILVAGLANGLIALLRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEDLTINQHR 60
 DB 1 MOPHYDLILVAGLANGLIALLRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEDLTINQHR 60
 QY 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
 DB 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHAFTVIDRGYTPDSALVGVQFQAFIGQEWQLSAPHGLSSPIIMDATVDOQ 180
 DB 121 SVRLKKGQVIGARAVIDGRGYAANSALS SVGFQAFIGQEWRLSHPHGLSSPIIMDATVDOQ 180
 QY 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNI RDYAAQGWPLQTLTLLREEQAL 240
 DB 181 NGYRFVYSLPSPTLLIEDTHYIDNATLDPECARQNI CDYAAQGWQLQTLTLLREEQAL 240
 QY 241 PITLTGDRNRQFWQQQPQACSGLRAGLPHFTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
 DB 241 PITLSGNADAFWQQRFLACSGLRAGLPHFTTGYSLPLAVALADRLSALDVFTSSVHAI 300
 QY 301 AHFAQORWQQQGFPRMLNRMFLAGPADSRNRVQRFYGLPEDIARFYAGKLTVTDLRL 360
 DB 301 THFARERWQQQGFPRMLNRMFLAGPADSRNRVQRFYGLPEDIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVLAALQAIMTTHR 382
 RESULT 6
 AA087891
 ID AA087891 standard; protein; 382 AA.
 XX AA087891;
 AC
 XX 17-OCT-2003 (revised)
 DT 10-MAR-1999 (first entry)
 XX

DE Protein encoded by the carotenoid biosynthesis gene crtY.
 XX Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene; crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside; carotenoid glucoside; crtY gene; crtZ gene; crtX gene; crtW gene; food additive.
 XX Pantoea ananatis.
 XX JP10327865-A.
 XX 15-DEC-1998.
 XX 29-MAY-1997; 97JP-00140460.
 XX 29-MAY-1997; 97JP-00140460.
 XX (KIRI) KIRIN BREWERY KK.
 XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
 XX WPI; 1999-099030/09.
 XX N-PSDB; AAV84083.
 XX New carotenoid glucoside(s) - used as food additives.
 XX Disclosure; Page 22-23; 26pp; Japanese.
 XX The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glucoside, in which all, or part of, carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)
 XX SQ Sequence 382 AA;
 Query Match 83.0%; Score 1678; DB 2; Length 382;
 Best Local Similarity 82.2%; Pred. No. 1e-157;
 Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;
 QY 1 MOPHYDLILVAGLANGLIALLRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEDLTINQHR 60
 DB 1 MOPHYDLILVAGLANGLIALLRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEDLTINQHR 60
 QY 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
 DB 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFGOHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHAFTVIDRGYTPDSALVGVQFQAFIGQEWQLSAPHGLSSPIIMDATVDOQ 180
 DB 121 SVRLKKGQVIGARAVIDGRGYAANSALS SVGFQAFIGQEWRLSHPHGLSSPIIMDATVDOQ 180
 QY 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNI RDYAAQGWPLQTLTLLREEQAL 240
 DB 181 NGYRFVYSLPSPTLLIEDTHYIDNATLDPECARQNI CDYAAQGWQLQTLTLLREEQAL 240
 QY 241 PITLTGDRNRQFWQQQPQACSGLRAGLPHFTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
 DB 241 PITLSGNADAFWQQRFLACSGLRAGLPHFTTGYSLPLAVALADRLSALDVFTSSVHAI 300
 QY 301 AHFAQORWQQQGFPRMLNRMFLAGPADSRNRVQRFYGLPEDIARFYAGKLTVTDLRL 360
 DB 301 THFARERWQQQGFPRMLNRMFLAGPADSRNRVQRFYGLPEDIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVLAALQAIMTTHR 382

RESULT 7

AW99100
AAW99100 standard; protein; 382 AA.

AAW99100;

17-OCT-2003 (revised)

14-MAY-1999 (first entry)

Erwinia uredovora crty protein sequence.

Beta-carotene hydroxylase; crty; crtB; crtE; crtI; xanthophyll;

metabolite.

Pantoea ananatis.

JP11046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

07-AUG-1997; 97JP-00213648.

(KIRI) KIRIN BREWERY KK.

WPI; 1999-208113/18.

N-PSDB; AAW19120.

Beta-carotin hydroxylase - useful for preparation of xanthophylls and their metabolites.

Disclosure; Page 15-17; 17pp; Japanese.

The present invention describes beta-carotin hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crty protein sequence from the present invention. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 382 AA;

Query Match 83.0%; Score 1678; DB 2; Length 382;

Best Local Similarity 82.2%; Pred. No. 1e-157;

Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

Y 1 MSHYDLILVGLAGLIALRLQQOHPDMRILLIENGPAGNHTWSPHEEDLTINQHR 60
b 1 MSHYDLILVGLAGLIALRLQQOHPDMRILLIENGPAGNHTWSPHEEDLTINQHR 60
Y 61 WIAPLVHHWPDYQVFPQRRHNSGYCVTSRHPAGILRQOFGQHLMLHTANSVRAE 120
b 61 WIAPLVHHWPDYQVFPQRRHNSGYCVTSRHPAGILRQOFGQHLMLHTANSVRAE 120
Y 121 SVOLADGRIIHASTVIDGRGYPDSALRVGFAFTGQEWQLSAPHLSPIMDATVDOO 180
b 121 SVRLKGVGIGARAVIDGRGYAANSALSIVGFAFTGQEWQLSAPHLSPIMDATVDOO 180
Y 181 NGYRFVYPLSATALIIEIDTHYIDKANLQAEARQNIIDYAAQGMPLQTLREEQAL 240
b 181 NGYRFVYPLSPTLLIEDTHYIDNATIDPECARQNICDYAAQGMPLQTLREEQAL 240
Y 241 PITLTGDNQFQWQOQACSLRAGLFHTTGYSLPLAVADRLSALDVFTSSVHOTI 300
b 241 PITLSGNADAFWQOQRLACSLRAGLFHTTGYSLPLAVADRLSALDVFTSSVHOTI 300
Y 301 AHPAQORWQOQGFPRMLNMLFLAGPABSRWRVMORFYGLPDLIARFYAGKLTIDRLR 360
b 301 THPARERWQOQGFPRMLNMLFLAGPABSRWRVMORFYGLPDLIARFYAGKLTIDRLR 360
Y 361 ILSGKPPVPVFAALQAIMTTHR 382
b 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 8

AAW01125

ID AAW01125 standard; protein; 374 AA.

XX AAW01125;

16-OCT-2003 (revised)

25-MAR-2003 (revised)

11-DEC-1996 (first entry)

Lycopene cyclase.

GPP synthase; lycopene cyclase; phytoene synthase; beta-carotene; phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside; pigment; food colourant; chloroplast transit peptide; increase yield; tobacco ribulose bis-phosphate carboxylase-oxygenase.

Pantoea agglomerans.

US5530188-A.

25-JUN-1996.

21-JUL-1993; 93US-00095726.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

30-OCT-1991; 91US-00785566.

(STAD) AMOCO CORP.

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;

Ausich RL;

WPI; 1996-308823/31.

N-PSDB; AAT40795, AAT40796.

Increasing prodn. of total carotenoid(s) in a higher plant - by transforming with vector encoding chloroplast transit peptide operably linked to the Erwinia herbicola lycopene cyclase structural gene.

Example 17; Col 101-104; 99pp; English.

Manipulation of the DNA encoding the present sequence, lycopene cyclase, by in frame linkage to the chloroplast transit peptide (AAW01124) of the tobacco ribulose bis-phosphate carboxylase-oxygenase gene can lead to increased production of total carotenoids in the chloroplast of transformed plants as compared to native, non-transformed plants of the same type. Beta-carotene is an effective and apparently harmless food colourant and is also in the pathway for biological synthesis of further C40 carotenoids such as zeaxanthin and zeaxanthin diglucoside. Other enzymes involved in the carotenoid biosynthesis pathway include geranylgeranyl pyrophosphate (AAW01119), phytoene synthase (W01121) and phytoene dehydrogenase-4H (AAW01123). (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 374 AA;

Query Match 55.8%; Score 1127; DB 2; Length 374;

Best Local Similarity 58.4%; Pred. No. 5.6e-103;

Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

QY 6 DLILVGLAGLIALRLQQOHPDMRILLIENGPAGNHTWSPHEEDLTINQHRWIAPL 65
b 3 DLILVGLAGLIALRLQQOHPDMRILLIENGPAGNHTWSPHEEDLTINQHRWIAPL 62
QY 66 WWHWPDYQVFPQRRHNSGYCVTSRHPAGILRQOFGQHLMLHTANSVRAESVQLA 125
b 63 VAHAWPGYEVQFPDLRRRLRGYISITSERFALHQAALGENIWLNCVSSEVLPNSVRLA 122

126 DGRTHASTVIDGRGTPDSALRVGFQAFIGQWQLSAPHGLSSPIIMDATVDQNGYRF 185
123 NGEALLAGAVIDGRGTVTASSAMQTGYQLFGQQWRLTQPHLTVPILMDATVAQQQGYRF 182
186 VYTLPLSATALLEDTHYIDKANLQAEARQNTDYAARQWPLQTLRLREOGALPITLT 245
183 VYTLPLSADTLLIEDTRYANVPQDDNALRQVTDYAHKSGWQLAQLEREETGCLPITWR 242
246 GDNRFQWQQPQACSGLRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 305
243 VTSRLCGEMRRRAASGRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 302
306 QRWQQQGFRLNMLFLAGPAESRWVQRFYGLPEDLIARFYAGKLVTDRLRLISGK 365
303 RHWRRQGFRLNMLFLAGREENRWVQRFYGLPEPTVERFYAGRLSLFDKARILITGK 362
366 PPVPV 370
363 PPVPL 367

RESULT 9
AAW32474
ID AAW32474 standard; protein; 374 AA.
XX
AC AAW32474;
XX
DT 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 15-JAN-1998 (first entry)
XX
DE Erwinia herbicola lycopene cyclase.
XX
KW Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;
KW lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;
KW yeast; plant; vitamin A; cancer.
XX
OS Pantoea agglomerans.
XX
FH Key
FT Active-site 1 Location/Qualifiers
FT /note= "Encoded by GTG in wild-type, but by ATG in the
FT genetically engineered form"
XX
PN US5656472-A.
XX
PD 12-AUG-1997.
XX
PF 07-JUN-1995; 95US-00473512.
XX
PR 02-MAR-1990; 90US-00487613.
PR 18-MAY-1990; 90US-00525551.
PR 03-AUG-1990; 90US-00582674.
PR 28-FEB-1991; 91US-00662921.
PR 21-JUL-1993; 93US-00095726.
XX
PA (STAD) AMOCO CORP.
XX
PI Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;
PI Brinkhaus FL;
XX
DR N-PSDB; AAT91547.
XX
PT DNA encoding Erwinia herbicola lycopene cyclase - for producing
PT recombinant enzyme, and transgenic organisms with increased beta-carotene
PT levels.
XX
PS Example 17; Fig 19; 102pp; English.
XX
SS A novel DNA molecule has been isolated which encodes an Erwinia herbicola
CC lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA

CC molecule comprises at least 1125 bp and is present in the plasmids
CC PARC147, PARC1509, PARC1510 and PARC1520. The present sequence represents
CC lycopene cyclase. The new DNA molecule can be used to produce the
CC recombinant enzyme and transgenic organisms, e.g. yeasts or plants, with
CC increased beta-carotene levels. Beta-carotene is used as a colourant in
CC margarine and butter and as an intermediate for vitamin A, and may
CC prevent cancer. (Updated on 25-MAR-2003 to correct pf field.) (Updated on
CC 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 374 AA;
Query Match 55.8%; Score 1127; DB 2; Length 374;
Best Local Similarity 58.4%; Pred. No. 5.6e-103;
Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;
QY 6 DLILVGLAGLIALPLQOQHDPMDRILLLEAGPEAGGNHTWSFHEDTLNQHRTAPL 65
DB 3 DLILVGGGLAGLIALRLRQYLPQNLILLLEAGPEAGGNHTWSFHEDDLTPGQHAWLAPL 62
QY 66 VVHEWPDYQVRFQRRRHVNSGYCYVTSRHFAGILRQCFQHLWLHTAVSAEASVQLA 125
DB 53 VAHWPFGVEVQFPDLRRRLARGYVITSERFAEALHQALGENIWLNCSEVLPSVRLA 122
QY 126 DGRTHASTVIDGRGTPDSALRVGFQAFIGQWQLSAPHGLSSPIIMDATVDQNGYRF 185
DB 123 NGEALLAGAVIDGRGTVTASSAMQTGYQLFGQQWRLTQPHLTVPILMDATVAQQQGYRF 182
QY 186 VYTLPLSATALLEDTHYIDKANLQAEARQNTDYAARQWPLQTLRLREOGALPITLT 245
DB 183 VYTLPLSADTLLIEDTRYANVPQDDNALRQVTDYAHKSGWQLAQLEREETGCLPITWR 242
QY 246 GDNRFQWQQPQACSGLRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 305
DB 243 VTSRLCGEMRRRAASGRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 302
QY 306 QRWQQQGFRLNMLFLAGPAESRWVQRFYGLPEDLIARFYAGKLVTDRLRLISGK 365
DB 303 RHWRRQGFRLNMLFLAGREENRWVQRFYGLPEPTVERFYAGRLSLFDKARILITGK 362
QY 366 PPVPV 370
DB 363 PPVPL 367
RESULT 10
AAR13987
ID AAR13987 standard; protein; 374 AA.
XX
AC AAR13987;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-NOV-1991 (first entry)
XX
DE Lycopene cyclase - variant.
XX
KW GGPP; carotenoid; phytoene; zeaxanthin; lycopene.
XX
OS Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).
XX
PN MO9113078-A.
XX
PD 05-SEP-1991.
XX
PF 04-MAR-1991; 91WO-US001458.
XX
PR 02-MAR-1990; 90US-00487613.
PR 18-MAY-1990; 90US-00525551.
PR 03-AUG-1990; 90US-00582674.
PR 28-FEB-1991; 91US-00662921.
XX
PA (STAD) AMOCO CORP.
XX

Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
 Yen HC;
 WPI; 1991-281410/38.
 R N-PSDB; AAQ13723.
 X
 X Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA
 T encoding enzymes from *Erwinia herbicola*.
 X
 X Disclosure; Fig 19(1-3); 31pp; English.
 X
 X There are a total of six relevant genes in a 7900 bp region that cause E.
 C coli cells to produce GGPP and the carotenoids phytoene through
 C zeaxanthin diglucoside, which is the final prod. identified in the
 C carotenoid pathway contd. in plasmid PARC376 (contg. a ca. 13 kb
 C chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607
 C (1986)). The genes for geranylgeranyl pyrophosphate (GGPP) synthase,
 C phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-
 C carotene hydroxylase, and zeaxanthin glycosylase are represented in
 C AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726
 C respectively. The native sequence (AAQ13722) encoding this protein was
 C genetically engineered for use in expression in yeast. At the 5' end of
 C the gene, the native initiation GTG codon has been changed to an ATG
 C codon. The second amino acid residue, Arg, was originally encoded by an
 C AGG codon that was changed to a CGG codon, while retaining its coding for
 C the Arg amino acid residue. Recombinant expression plasmids can be used
 C to produce large amts. of the enzymes and hence large amts. of the
 C carotenoids which they synthesise. (Updated on 25-MAR-2003 to correct PP
 C field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 24-OCT-
 C 2003 to standardise OS field)
 X
 X Sequence 374 AA;
 X

Query Match	55.7%;	Score	1126;	DB	2;	Length	374;
Best Local Similarity	58.4%;	Pred. No.	7e-103;				
Matches	213;	Conservative	53;	Mismatches	99;	Indels	0;
2Y	6	DLILVAGLGLIALRLQOQHDPDMRILLIEAGPEAGGNHTWSFHEEDLTLNQHRWTA	165				
2b	3	DLILVGRGLANGLIANLRQRYFQILNLLIEAGQFCGNHTWSFHEDDLTFCQHANLAPL	62				
2Y	66	VVHWPPQYVFPQRRRHVNSGYCVTSRHFAGILRQFGQHLHLHTAVSAVHAESVOLA	125				
2b	63	VAHAWPGYEVQFPLDRLRLRARGVYSITSERFAEALHQALGENIWLNCSEVLPNSVELA	122				
2Y	126	DGRIIHASTVTDIGRGYTFDSALNRVQAFIQGSEWQLSAPHEGLSSPTIMDATVDQONGYR	185				
2b	123	NGEALLGAVIDGRGVTVASSAMQTGYFLFGQWRLTQPHGLTVPILMDATVAAQQGGRF	182				
2Y	186	VYTPLSATALLIEDTHYIDKANLQAEARQNIIRDYAARGWPLQTLLEEGCALPITET	245				
2b	183	VYTPLSADTLIEDTRYANVPQRDNNALRQITDYAHSKGMQLAQLEREETGCLPTWR	242				
2Y	246	GDNRQFWMQOQACGLRAGLFHPTTGYSLPLAVALADRLSALDVFPTSSSVHQTIAHQAQ	305				
2b	243	VTSRLCGPMRSPRAASGRAGLFHPTTGYSLPLAVALADALADSFRUGSVPLLYQLTRQFAE	302				
2Y	306	ORWOOQGFRLINMLFLAGPAESRWRVMORFYGLPEDIARFYAGKLVTTVTRLRLISGK	365				
2b	303	RHWRRQGFRLINMLFLAGEENRWRVMORFYGLPEPTVERFYAGRLSLFPDKARILTGK	362				
2Y	366	ppvvpv	370				
2b	363	ppvpl	367				

```

RESULT 11
ABM70122
ID   ABM70122 standard; protein; 392 AA.
XX
XX   ABM70122;
AC   AC
XX
XX   20-NOV-2003 (first entry)
DT

```

XX	Photorhabdus luminescens protein sequence #3219.
XX	
XX	
KW	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW	Detection; food; gene expression; plant; animal; microorganism; toxin;
KW	antibiotic; biopesticide; virulence factor; disease model; plague;
KW	whooping cough.
XX	
XX	
XX	Photorhabdus luminescens.
OS	
XX	
PN	WO200294867-A2.
XX	
XX	29-NOV-2002.
XX	
XX	07-FEB-2002; 2002WO-IB003040.
XX	
XX	07-FEB-2001; 2001FR-00001659.
PR	
XX	(INSP) INST PASTEUR.
XX	(CNRS) CNRS CENT NAT RECH SCI.
PA	
XX	Duchaud B, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI	Buchrieser C;
PI	
XX	
XX	WPI; 2003-148459/14.
XX	
XX	
PT	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX	
XX	Claim 2; SEQ ID NO 3219; 1205pp; French.
PS	

The invention relates to the isolation of genes and their encoded proteins from *Photobacterium luminescens*. The isolated sequences are sources of probes and primers for detecting the genome of *P. luminescens* and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than *P. luminescens* and are able to alter response or sensitivity to toxins and antibiotics produced by *P. luminescens*. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibiotics useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which *P. luminescens* is a model (particularly plague and whooping cough). This sequence represents one of the isolated *P. luminescens* proteins

Sequence 392 AA;

Query Match	51.5%;	Score 1040.5;	DB 6;	Length 392;
Best Local Similarity	52.5%;	Pred. No. 2.4e-94;		
Matches 201;	Conservative	51;	Mismatches 130;	Indels 1;
Gaps	1;			
Qy	1	MQHYDILVAGLANGLIALRQQQHPDMRILLIRAGPAGGNGHTWSPHEEDITLQHR	60	
Db	1	MMYDWDILVGGGLANGLIARFQCCPHLRVLLIENETETGGNHTWSPHQDUEAHE	60	
Qy	61	WIAVLVHHFPDYOVREPORRHNSGYCYVTSRHPAGILRQQFGQHLWHTAVSAVAAE	120	
Db	61	WIAPLTYRMSGYDVIFFAPQRTLPHPYSFYTTSQHPASIIHVALGERIQITLLVORLTPQ	120	
Qy	121	SVQLADGRITHASVTDIGRVTPDSALRVGFGAPFGQEWOLSPAGHGLSSPTIMDATVDQO	180	
Db	121	KYVLQDSSLSAGAVIDGRGWRPGPFICSGTQAPFGQEWELSEHSHTPLIMBTSVQOD	180	
Qy	181	NGYRFVTLPLSATAHLEIETHYIDKANLQAEARONIRDYARQGWPLQTLLEEQCAL	240	

Db 181 TGYRIFYVLPSSRLLIEDTHYVDGPPDVALSQATIAEYAKKHGKGLRESGCL 240
QY 241 PITLTGDRQFQQ-QPQACSGLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSVHQT 299
Db 241 PITLTGDRQFQQ-QPQACSGLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSVHQT 300
QY 300 IAHFAQQWQOQGFRLNMLFLAGPAESRWVQRFYGLPEDLIARFVAGKLTVDRL 359
Db 301 LRDYAQWQWQRFRLNMLFLAGDQQRWQWQRFYGLPEDLIARFVAGKLTVDRL 360
QY 360 RILSGKPPVPVFAALQAI 372
Db 361 RILSGKPPVPVFAALQAI 372
RESULT 12
ID AAW06517 standard; protein; 382 AA.
XX AAW06517;
XX 17-OCT-2003 (revised)
DT 08-MAR-1997 (first entry)
XX Flavobacterium lycopene cyclase.
XX Carotenoid; lycopene; beta-carotene; echinenone; canthaxanthin;
XX zeaxanthin; adonixanthin; astaxanthin; crtY; lycopene synthase.
XX Flavobacterium sp. ATCC 21588; WT (ATCC 21588).
XX BP747483-A2.
XX 11-DEC-1996.
XX 29-MAY-1996; 95EP-00108556.
XX 09-JUN-1995; 95EP-00108888.
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX Hohmann H, Pasamontes L, Tessier M, Van Loon A;
XX WPI; 1997-023160/93.
XX N-PSDB; AAW45143.
XX Flavobacterium gene sequences encoding carotenoid biosynthesis enzymes -
XX for the production of carotenoid(s), useful in foods and animal feeds.
XX Example 2; Fig 7; 80pp; English.
XX Lycopene synthase (AAW06517) is the product of the crtY gene identified
XX in a carotenoid gene cluster (see also AAW45143) of Flavobacterium sp.
XX R1534. This 42368 Da protein introduces the beta-ionone rings at both
XX sides of lycopene to obtain beta- carotene. Enzymes of the Flavobacterium
XX carotenoid biosynthetic pathway (see also AAW06515-18 and AAW00871) can
XX be expressed, optionally with Alcaligenes beta-carotene beta-4-oxigenase,
XX in host cells for the prodn. of lycopene, astaxanthin, beta-carotene,
XX echinenone, zeaxanthin, canthaxanthin and adonixanthin. (Updated on 17-
XX OCT-2003 to standardise OS field)
XX Sequence 382 AA;
Query Match 38.8%; Score 783.5; DB 2; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.9e-69;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 5;
QY 5 YDILLVAGLANGLIALRLQOQPMRILLIEAGPEAGQNHNTWSFHEEDLTINQHWIA- 63
Db 3 HDLLIAGLUGSLIALAVDRRPDARIWLDARGSPSQHTWSCHTDLS---PEMLAR 59
QY 64 --PLVGHHPDVQVRFQRRHVNSSYCYVTSSEHAGILRQGFQHLMLHTAVSAVHBS 121

Db 60 LSPIRRGEWTQDEVAFPDHSRRLTGTGYSIEAGALIGLLQ---GVDLRWNTHVATLDDTG 116
QY 122 VOLADGRIIHAFTVIDRGCTTPDSALVGFQAFIGQWQLSAPHLSSPIIMDATVDQON 181
Db 117 ATLTDGSIIEAACVIDARGAVETPHLTVGFKFVGVEIETDAPHGVERPMINDATVPQMD 176
QY 182 GYRFVYVTLPLSATALLEDTHYVDKMLQAEFRARQNIQDYAARQGWPLQTLREEQCALP 241
Db 177 GYRFYLLFPFPTLIEDTRYSDGDLDDGALAQAASLDYAAREGWTGQE-MRERGIPL 235
QY 242 IELTDGNEQFQQOPOAC--SGLRAGLPHPTTGYSLPLAVALADRLSALDVFTSSVHQT 299
Db 236 IALAHDALIGFWRDHAQGAVPVGLAGLPHPTTGYSLPYAAQVADAIARDL-TTASARRA 294
QY 300 IAHFAQQWQOQGFRLNMLFLAGPAESRWVQRFYGLPEDLIARFVAGKLTVDRL 359
Db 295 VEGWALDRAADRFLRLNMLFRGCPDREYRLQRFYRLPQLIERFYAGRLTLADRL 354
QY 360 RILSGKPPVPVFAALQAI 377
Db 355 RIVTGRPPILSQAVRCL 372
RESULT 13
ID AAW69533 standard; protein; 382 AA.
XX AAW69533;
XX 10-AUG-1999 (first entry)
XX Flavobacterium sp. R1534 crtY gene product lycopene cyclase.
XX Carotenoid; pigment; canthaxanthin; R1534; crtB; crtB; prephytoene synthase;
XX crtI; phytoene desaturase; crtY; lycopene cyclase; GGPP synthase; crtB;
XX crtW2396; beta-carotene beta-oxygenase; food product; fermentation.
XX Flavobacterium sp.
XX JP10155497-A.
XX 16-JUN-1998.
XX 02-DEC-1997; 97JP-00348653.
XX 02-DEC-1996; 96EP-00810839.
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX WPI; 1998-391048/34.
XX N-PSDB; AAW40146.
XX Preparation of carotenoid - comprises fermentation with transformed cell.
XX Claim 1; Fig 25; 80pp; Japanese.
XX The invention describes the preparation of carotenoid pigments e.g.
XX canthaxanthins using a cell transformed by a vector having DNA sequences
XX (a) to (e) or substantially homologous sequences. (a) a DNA sequence
XX (crtB) coding GGPP synthase of Flavobacterium sp. R1534; (b) a DNA
XX sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534;
XX (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium
XX sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of
XX Flavobacterium sp. R1534, and (e) a DNA sequence (crtW2396) coding beta-
XX carotene beta-oxygenase of a microbe E-396 (FERM BP-4283). The carotenoid
XX or a carotenoid mixture can also be used in preparation of food products.
XX The method is an improved method of fermentation for carotenoid
XX production
XX Sequence 382 AA;
Query Match 38.8%; Score 783.5; DB 2; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.9e-69;

Matches	165;	Conservative	64;	Mismatches	136;	Indels	13;	Gaps	6;				
y	5	YDILVLGAGL	GLIALRLQ	QQHPDMRILLISAGPEAGCNHTWS	HEEDLTINCRWIA-	63							
b	3	HDLIAGAGLS	GALIALAVDR	RPDPARIYVLD	ARSQPSQHTWSCHD	TLS---	PEWLAR	59					
y	64	--PLVVHHPD	QVVRPQRRR	HVNSGYCVTSRHFAGILRQ	QFGQHLMLHTAVASV	HAES	121						
b	60	LSPIRRGENT	QDQVAFPD	HSRLRTTGYGIEBAGALIGLLQ---	GYDLEWNTHTVLD	DTG	116						
y	122	VQLADGRITH	ASVVIDGR	GYTPD	SALRVGFQAFIGQEWCLSPHGLSSP	IIMDATV	DQON	181					
b	117	ATUTDSRIE	ACVLDARG	AVETPLTVGQFQVGVIE	ETDPHGV	ERPMDATV	PQMD	176					
y	182	GYRFVYTL	PUSAFAL	IEBOTHYIDKANLQ	AERANIRDYAARQGW	PLQTLIR	EEQALP	241					
b	177	GYRFIVLL	PPSPTRIL	IEDTRYSDG	EDLDGALAQASLDY	AARQWTCQ	Q--KR	ERGTLIP	235				
y	242	ITLTGNRQ	FQWQOPAC--	SGLRAGLFHPT	TGYSLPLAVAL	ABELSALD	VFTSS	VHOT	299				
b	236	IALAHA	DAIGFWRD	HAQGV	VPVGLG	LFHVTGYSLP	VAAQVADA	IAARDL--TT	ASARRA	294			
y	300	IAHFAQRW	QOQQGF	FRLNRLFL	AGP	ASSRWK	VPVGLP	PEDILAF	RYAGKLT	TVTDRL	359		
b	295	VRGWA	LD	RADRD	RFLRLN	RLFRG	CP	DRRYLL	QRFV	LPPLIER	PYAGRLT	ADRL	354
y	360	RILSGK	PPVPVFA	ALQAI	377								
y	355	RIVTGR	PPPLSQ	AVRCL	372								

ESULT 14
AR79062
D AAR79062 standard; peptide; 386 AA.

A		
X	AAR79062;	
C		
X		
T	27-AUG-2003	{revised}
T	28-FEB-1996	{first entry}

E 3 hydroxy-beta-ionone ring methylene to keto group converting peptide.

Xanthophyll¹: astaxanthine; methylene; keto group; conversion;

3-hydroxy-beta-ionone ring.

45 *Agrobacterium aurantiacum.*

X N WO9518220-A1.

X D 06-III.-1995

26-DEC-1994: 94W0-TP002220

X 27-DEC-1993. 93.TP-00348737

27-DEC-1993; 93JP-00348737.
05-SEP-1994; 94JP-00235917.

X A (KTST) KTRTN BEER KK

A (KIRI) KIRIN BEER KK.
A (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.

X T Misawa N. Kondo K. Kajiwara S. Yokoyama A:

X B WPT: 1995-246386/32

R R
R B

DNA's encoding xanthophyll(s) - esp. asta:xanthin and other xanthophyll(s) using e.g. E. coli.

Disclosure: Page 21: 131pp: Japanese.

AA79058-R790629 are xanthophyll polypeptides. These polypeptides are capable of converting the 4-methylene group of a 3-hydroxy-beta-ionone ring to a 4-keto group in doing so these peptides also add a hydroxyl group to the 3-position carbon-atom of the 4-keto-beta-ionone ring. The

CC DNA sequences may be used in the production of astaxanthine and other
CC keto gp. contg. xanthophylls, the sequences may also be used to transform
CC certain yeasts and other microorganisms. (Updated on 27-AUG-2003 to
CC correct OS field.)
CC
XX
SO Sequence 386 AA;

Query Match	37.6%	Score 759.5;	DB 2;	Length 386;
Best Local Similarity	43.0%;	Pred. No. 2e-66;		
Matches 164:	Conservative	65;	Mismatches 137;	Indels 15;
				Gaps 8;

[illegible]

RESULT 15

AAW87884
ID AAW87884 standard; protein; 386 AA.

XX AC BAW87884:

XX
DT 10-MAR-1999 (first entry)

XX DE protein encoded by the carotenoid biosynthesis gene crty.

XX Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;
KW
KW crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;
KW carotenoid glycoside; crtY gene; crtZ gene; crtW gene;
KW food additive.

XX OS *Agrobacterium aurantiacum*.XX
PN
TP10327865-AXX
15-DEC-1998
PDXX
DE 29-MAY-1997. 97JP-00140460.

XX 29-MAY-1997. 97JP-00140460

XX
DA (ZTPT 1 KTPTN BREWERY KK

PA (KIRI) KIRIN BREWERY KK.
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

XX
WP: 1999-099030/09

DR N-PSDB; AAV87884.
YY

New carotenoid glucoside(s) - used as food additives.

pt New carotenoid glucoside(s) - used as food additives.
xx

X

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:33:49 ; Search time 35.4397 seconds
(without alignments)
3837.172 Million cell updates/sec

file: US-09-941-947A-28

effect score: 2231
sequence: 1 MSHFAVIAPPPFHVRLQN.....EQAMRTQPVLSGGDYATAT 431

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

otal number of hits satisfying chosen parameters: 1017041

inimum DB seq length: 0

Maximum Match 100%

Maximum DB seq length: 2000000000

Maximum Match 100%

Listing first 45 summaries

atabase :

SPREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2231	100.0	431	2	Q8GCS3
2	1794	80.4	431	2	Q47843
3	1579.5	70.8	431	2	Q8VUJ7
4	558.5	25.0	419	16	P72650
5	262	11.7	397	16	Q9S1V2
6	251.5	11.3	397	16	Q825U1
7	247	11.1	412	16	Q9S0P6
8	243.5	10.9	389	2	Q83X61
9	239.5	10.7	461	2	Q9XC67
10	238.5	10.7	402	16	Q81E86
11	234.5	10.5	402	16	Q81R94
12	228.5	10.2	389	16	Q8DSH9
13	219.5	9.8	379	2	Q93H13
14	219	9.8	417	2	Q86304
15	217	9.7	424	2	Q68841
16	213	9.5	446	2	Q83WG5

17	210.5	9.4	427	16	Q98SL9
18	207	9.3	397	16	Q81CG5
19	207	9.3	406	16	Q97TQ3
20	204	9.1	392	16	Q34539
21	201	9.0	397	2	Q8KNE0
22	198.5	8.9	418	16	Q9RYI3
23	197.5	8.9	418	16	Q9ADH3
24	197	8.8	450	16	Q89RS0
25	195.5	8.8	287	16	Q31853
26	187.5	8.4	421	2	Q8KHE4
27	181.5	8.1	540	5	Q9VCL3
28	178.5	8.0	402	2	Q8KHU5
29	178.5	8.0	492	10	Q9SBL1
30	177	7.9	594	5	Q16243
31	175.5	7.9	414	2	Q98595
32	175	7.8	407	16	Q97FM0
33	175	7.8	418	16	Q8PJG6
34	172.5	7.7	530	11	Q9R110
35	172	7.7	388	16	Q7Y01
36	172	7.7	420	16	Q33282
37	171	7.7	382	2	Q9F2F9
38	170.5	7.6	401	2	Q8KNC3
39	169.5	7.6	333	16	Q81CW9
40	167	7.5	422	2	Q70023
41	164.5	7.4	388	2	Q9REAL
42	164.5	7.4	406	2	Q33MW2
43	164.5	7.4	541	11	Q91WS7
44	164	7.4	392	16	Q49841
45	163	7.3	529	11	Q8VIF8
					Q98SL9 rhizobium 1
					Q81CG5 bacillus ce
					Q97TQ3 clostridium
					Q34539 bacillus su
					Q8KNE0 micromonosp
					Q9RYI3 deinococcus
					Q9ADH3 streptomyces
					Q89RS0 bradyrhizob
					Q31853 bacillus su
					Q8KHE4 nocardia ae
					Q9VCL3 drosophila
					Q8KHU5 actinosynne
					Q9SBL1 sorghum bic
					Q16243 caenorhabdi
					Q98595 nocardia ae
					Q97FM0 clostridium
					Q8PJG6 xanthomonas
					Q9R110 cavia porce
					Q7Y01 mycobacteri
					Q33282 mycobacteri
					Q9F2F9 streptomyce
					Q8KNC3 micromonosp
					Q81CW9 bacillus ce
					Q70023 streptomyce
					Q9REAL streptomyce
					Q33MW2 nocardia br
					Q91WS7 mus musculu
					Q49841 mycobacteri
					Q8VIF8 cavia porce

ALIGNMENTS

RESULT 1

ID Q8GCS3 PRELIMINARY; PRT; 431 AA.
AC Q8GCS3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zeaxanthin glucosyl transferase.
GN CRTX.

OS Pantoea stewartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=66269;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 8200;
RA deSouza M.L., Kollmann S.R., Schroeder W.A.;

RT "Carotenoid Biosynthesis (WO 02/079395 A2).";

RL Submitted (OCT-2002) to the EMBL/GenBank/DDAJ databases.

DR EMBL; AI166713; AAN85597.1; -

DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR006326; UDPGT_MGT.

DR InterPro; IPR002213; UDP_gluco_trans.

DR Pfam; PF00201; UDRGT, 1.

DR TIGRFAMs; TIGR01426; MGT; 1.

DR Transferase.

SW SEQUENCE 431 AA; 46611 MW; A9BC082567039732 CRC64;

Query Match 100.0%; Score 2231; DB 2; Length 431;

Best Local Similarity 100.0%; Pred. No. 4.2e-159;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSHFAVIAPPPFHVRLQNLAQELVARGHVRVTPPQQHDCALVTGSDIGQTUCLQTHP 60

Db 1 MSHFAVIAPPPFHVRLQNLAQELVARGHVRVTPPQQHDCALVTGSDIGQTUCLQTHP 60

QY 61 PGSLSHLHLAAHPGPSMLRLINEMARTSDMLCRLPAFAHALQIEGVIVDQMEPAGAV 120

```
1b 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
1y 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180
1b 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180
1y 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDPFKALPDCCHAVGLPQPGTSS 240
1b 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDPFKALPDCCHAVGLPQPGTSS 240
1y 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
1b 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
1y 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
1b 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
1y 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
1b 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
1y 421 VLSGQDYATAT 431
1b 421 VLSGQDYATAT 431

RESULT 2
247843
ID Q47843 PRELIMINARY; PRT; 431 AA.
AC Q47843;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CrtX.
EN CrtX.
RN CrtX.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RX [1]
SQ SEQUENCE FROM N.A.
MEDLINE=94236237; PubMed=8180698;
RA TO K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
RA Chang Y.S., Liu S.T.;
RT "Analysis of the gene cluster encoding carotenoid biosynthesis in
RT Erwinia herbicola Eh013.";
RL Microbiology 140:331-339(1994).
DR EMBL; M90698; AAA21261.2; -.
DR PIR; S52583;
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 431 AA; 47070 MW; ED08B121D77C19CE CRC64;

Query Match 80.4%; Score 1794; DB 2; Length 431;
Best Local Similarity 82.4%; Pred. No. 2.5e-126;
Matches 355; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

QY 1 MSHFAVIAPPFFSHVRALONLAQELVARGHRTVFPQGHCKALVTGSDIGFQTVGLQTHP 60
Db 1 MSHFAIAAPPFFSHVRALONLAQELVARGHRTVFPQGHCKALVTGSDIGFQTVGLQTHP 60

QY 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
Db 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120

QY 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180
Db 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180

QY 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
Db 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300

QY 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
Db 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360

QY 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
Db 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420

QY 421 VLSGQDYATAT 431
Db 421 VLSGQDYATAT 431
```

```
QY 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDPFKALPDCCHAVGLPQPGTSS 240
Db 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDPFKALPDCCHAVGLPQPGTSS 240
QY 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
Db 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
QY 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
Db 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
QY 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
Db 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
QY 421 VLSGQDYATAT 431
Db 421 VLSGQDYATAT 431

RESULT 3
Q8VUU7
ID Q8VUU7 PRELIMINARY; PRT; 431 AA.
AC Q8VUU7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CrtX protein.
EN CrtX.
RN CrtX.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=182454;
RX [1]
SQ SEQUENCE FROM N.A.
RA Kamiunten H. Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79601.1; -.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 431 AA; 47439 MW; F7E124F2632A8EF7 CRC64;

Query Match 70.8%; Score 1579.5; DB 2; Length 431;
Best Local Similarity 70.9%; Pred. No. 3e-110;
Matches 304; Conservative 53; Mismatches 71; Indels 1; Gaps 1;

QY 1 MSHFAVIAPPFFSHVRALONLAQELVARGHRTVFPQGHCKALVTGSDIGFQTVGLQTHP 60
Db 1 MSHFAVIAPPFFSHVRALONLAQELVARGHRTVFPQGHCKALVTGSDIGFQTVGLQTHP 60

QY 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
Db 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120

QY 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180
Db 121 VASASGLPFFVSACALPLNREPGLPLAVMPFFYGTSDAAREYTTSEKIYDWMRHRDV 180

QY 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300
Db 241 TSYPSPDXKPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARG 300

QY 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360
Db 301 GDIQVDFADQSAALSOAQLTTHGGNMTVLDAISRTPLALPLAFDQPGVASRIVYHG 360

QY 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420
Db 361 IGKRASRFTTSHALAQIRSLTNTDYPQRMTKIOAALRLAGGTPAAADIIVEQAMRTCP 420

QY 421 VLSGQDYATAT 431
Db 421 VLSGQDYATAT 431
```



```
300 GGDIOVDFADQSAALSOAQLTIHGGKNTVLDALASRTPLALPLAFDQGVASRIYVH 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 ASHVQVDFADQAAALADAVITHGGKNTVLDGINHLTPTLPIAFDQGVAAVYVH 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 GIGKASFTTSHALARQIRSLTNTDYPQRTKIQAAELAGGTPAAADIVQAMETCQ 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 GIGKASFTTSHMARQQLTLLADESVQQMKTIRLSALRQAGGTTILAAIDIVEQAMLTRQ 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 PVLSQDYA 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 PVLTRHYA 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
P2650 PRELIMINARY; PRT; 419 AA.
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DB Zeaxanthin glucosyl transferase.
GN CRTX OR SLR1125.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasanoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nariu K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yabuta M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RA EMBL; D90899; BAA1652.1; -.
RA PIR; S74500; S74500.
GO GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
GO GO:0016740; P:transferase activity; IEA.
GO GO:0008152; P:metabolism; IEA.
GO GO:0006326; UDPGlcT.MGT.
InterPro: IPR002213; UDP_gluco_trans.
Pfam: PF00201; UDPGlcT.
TIGRFAMs: TIGR01426; MGT; 1.
Transferase; Complete proteome.
SEQUENCE 419 AA; 45330 MW; BAE7DFDE2592CB3 CRC64;

Query Match 25.0%; Score 558.5; DB 16; Length 419;
Best Local Similarity 31.7%; Pred. No. 1.1e-33;
Matches 132; Conservative 81; Mismatches 196; Indels 7; Gaps 5;

Y 1 MSHFAVIAPFFSHVRALQNLAEVLVARGHRTFFQHDCKALVTGSDIGFTVGLQTHP 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 1 MTHFGLCPATTGHLNTWMLPLCKELQORHTVTMGVLDAQAKTLAAGLNQAIATTEFP 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 61 PGSLSHLHLAHPGLPSMLR-LINEMARTSMCLCELPAPFALQIIEGVLVDQWEPAGA 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 61 LGAQAEFMAELKGLSGIKALQTVAKITQAKAAAFEEAPGVNAKAGVEVLVDQVQSEG 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 120 VVAEAGLFPVSVACALPNRPGPLVAMPPEYGTDAABERYTTSKIDYMLMRHHR 179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 121 TIGDRLGIFPISCSAVLNRRFTIPPATPWYDPSWGLQNLNGLNLRATKPTITA 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 180 VTAHACRMGLAPREKHLHCPSLAQISLIPELDPFKALPDCHVAGVPLRQCGTPGS 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 181 LINDYRQRNWLPAQSPNDYRSPYLAQISQQAAPFPPRECLPSRPHFTGTPFHSNVGRDIA 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 240 STSYFSPKPRIFASIGTLOGHRYGLFTTIKACEVDQALLAHCGLSATQAGELAR 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 241 DFPWEQLTQPIIYASLGHTIQNLQMSFTKILAEACMDLDAQLIIS-LGAKLESFALP- 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 300 GGDIOVDFADQSAALSOAQLTIHGGKNTVLDALASRTPLALPLAFDQGVASRIYVH 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 299 -GNPLVNYVAQLEHLQRTALFIHAGLTNTECLNNAVPMVAIPANDQGVAAIAMA 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 GIGK-RASRTTSHALARQIRSLTNTDYPQRTKIQAAELAGGTPAAADIVQEQ 413
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 358 GVGEFIPLSKLNTNN-LRAALEKVLTDTSYKNTLQLOQAIAKTAGGLTKAADIIEQ 412
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
Q9S1V2 PRELIMINARY; PRT; 407 AA.
ID Q9S1V2
AC Q9S1V2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DB Putative glycosyl transferase.
GN SCO0040 OR SCJ4.21.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
RA EMBL; AL939104; CAB52955.1; -.
DR PIR; T37104; T37104.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam: PF00201; UDPGlcT.
DR TIGRFAMs: TIGR01426; MGT; 1.
RX Transferase; Complete proteome.
SQ SEQUENCE 407 AA; 43498 MW; A61E94CF6088B4A0 CRC64;

Query Match 11.7%; Score 262; DB 16; Length 407;
Best Local Similarity 26.8%; Pred. No. 1.8e-11;
Matches 123; Conservative 59; Mismatches 165; Indels 112; Gaps 18;
```

2Y 1 MSHFAVIAPFFSHVRALQNLQALVARGHYTF--FQCHDCKALVTGSDIGFQTVGLQT 58
1 MSLAFINIGHGHIPTLPVLAELVRGHTVYHTFPAPFREIAATGANV-----RL 53
59 HPPGSLHLHLAAHPL--GPSMLRLINEMKARTSKMLCRLPAAFPALQIEG---VIVDQ 113
54 YPGGD-----QPLPDPAPITIMEALARTSLDL---LPAVLADLRDORPOLIVHDS 101
114 MEPAGVAEASGLPFFSVACALPLNREPGLPLAVMPPEYGTSDAARERTTSEKIYDML 173
102 ACPCGALAAARVLGLPAVSFTTAYNRHVSP-----TRASRE----- 139
174 MRHEDRVIAHACRMGLAPREKHLHCFSPALQISQ-----IPELDFPKALPDCHAVGP 229
140 -----LLAGAAAR-----PRLAGYVGARLALRRRFAATGVPLVD-----LAD 177
230 LROFGTGGSTSVFSP-----PKXPR-----IPASIGTLQ 261
178 IRQPLNLVYTSRAFOVAEVEFDRSYRVPFGSICARPDPSFFVNLRLDPVLVSLGTVEN 237
262 HRYGLPFTIAKACEEVDQAQLLHACGGLSATQAGELARGDIOVDFADQSAALSQAQLT 321
238 ADPLLETFATLSPL-AGVVVYSTGTQTPAALGELP--GNVLAERSVPQLEVLDRALF 294
322 ITGCGMNTVLDAISRTPLALPLATDQCVASRIVHGIGK--BASFT--TSHALAQ 377
295 ITGCGMNSVNEALFAGVPLLLVPGQADQPMVAERVVGLGAGLSITENTEDSVRAVAR 354
378 IRSLLTNTDYPQMTKIQALRLAGTTPAAADIVEQMR 416
355 ---LLEDSRYRAATSGLRATQHEAGGYRRAADELEGYLR 390
RESULT 6
Q825U1 PRELIMINARY; PRT; 397 AA.
ID Q825U1
AC Q825U1
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative glycosyl transferase.
GN SAV7358.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12592562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP065050; BAC75069.1; --
DR GO: GO:0016740; F:transferase activity; IEA.
DR Interpro: IPR006326; ODPRT_MGT.
DR TIGRPFAM: TIGR01426; MGT; I.
DR Transferase; Complete proteome.
SQ SEQUENCE 397 AA; 40868 MW; 77065B1CC04D6FE7 CRC64;

Query Match 11.3%; Score 251.5; DS 16; Length 397;
Best Local Similarity 26.6%; Pred. No. 1.1e-10;
Matches 115; Conservative 61; Mismatches 186; Indels 71; Gaps 18;
QY 1 MSHFAVIAPFFSHVRALQNLQALVARGHYTF--FQCHDCKALVTGSDIGFQTVGLQTHP 60
DB 1 MGRFVTVVPLVGVNPAVGTAAALAAAGHDIAWAGHPVLVRGLAGADAVVFFCAL---P 57
QY 61 PGSLHLHLAAHPLGPSMLRLINE--MARTSMLCRLPAAFPALQIEGVIVDQMEPAG 118
DB 58 EQLSRPAGLK---GPAFOFLWESFLVPLADAMAPGVRAAIEAYDPVVCDCQAVAG 113
QY 119 AVVASAGSLPFFSVACALPLNREPGLPLAVMPPEYGTSDAARERTTSEKIYDMLRRHD 178
DB 114 ALVASESLGRPAVTGATT---SAELVDPLAGMP-----KVAANL---D 149
QY 179 RVIAHACRM---GLA--PREKHLHCFSPALQISQISOLIPELDFPKALPDCHAVGP--LR 231
DB 150 GLLGLRLRRITGGAGADPR-----PSPGVLAYTTRALLGPVE-LPDRVVLVGPSVAA 202
QY 232 QPQGTGSGSTSVFSPDKPRIFASLGTLOGHRVGLFRTIAKACEEVDQAQLLHACGGLS- 290
DB 203 RPAGDPDPFMEWLEASALPTVLVSLGTANNDAGARFLNAA-----ASEAL-----GGIAD 251
QY 291 ATOAGELARGGDIO-----VVDFADQSAALSQAQLTITHCGMNTVLDAISRTPLA 342
DB 252 RYRAVLVDPGVVHPDPDTVLVRYVPLALLERLDVAVVCHAGNTVCEALHGVPLVV 311
QY 343 LPLADPQCVASRIVYHIGIKRASRTTSHA--LARQIRSLTNTDYPQMTK-IQAALR 399
DB 312 APIEDDQPIVAQVVDAGAGVRL-RFGRAAARIGAAVEAVLDPQAQHRKAAEAVGESFR 370
QY 400 LAGGTTPAAADIVE 412
DB 371 RAGGSESAADRL 383
RESULT 7
Q850P6 PRELIMINARY; PRT; 412 AA.
ID Q850P6
AC Q850P6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Glycosyl transferase.
GN AVEHI OR SAV945.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN SEQUENCE FROM N.A.
RX MEDLINE=9980548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macroide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12592562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

381 -LLNTDYPQMTKIOAALRAGGTAAADIVEQAM 415
353 ELTADEPVRSDALRAVRAEGTTRADIIIESAL 388

RESULT 9

99X67
ID Q9XC67 PRELIMINARY; PRT; 461 AA.
AC Q9XC67;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.-).
GN BC2066
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=226900;
XN TILCV.
XN Streptomyces fradiae.
XN Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
XN Streptomycineae; Streptomycetaceae; Streptomyces.
XN NCBI_TaxID=1906;
XN [1]
XN SEQUENCE FROM N.A.
XN STRAIN-T59235;
XN MEDLINE=50121747; PubMed=10458660;
XN Bate N., Butler A.R., Smith I.P., Cundliffe E.;
XN "The mycosae-biosynthetic genes of Streptomyces fradiae, producer of
XN tylosin.";
XN Microbiology 146:139-146 (2000).
XN EMBL; AF147704; AAD41824.1;
XN GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
XN GO; GO:0016740; P:transferase activity; IEA.
XN GO; GO:0005975; P:carbohydrate metabolism; IEA.
XN GO; GO:0003259; P:lipid glycosylation; IEA.
XN InterPro; IPR004276; Glyco_trans_28.
XN InterPro; IPR006326; UDPGT_MGT.
XN Pfam; PF03033; Glyco transf 28; 1.
XN TIGRFAMs; TIGR01426; MGT; 1.
XN Transferase.
XN SEQUENCE 461 AA; 50120 MW; 4A61AE1F06A2F59 CRC64;

Query Match 10.7%; Score 239.5; DB 2; Length 461;
Best Local Similarity 23.3%; Pred. No. 1.1e-09;
Matches 108; Conservative 58; Mismatches 179; Indels 119; Gaps 16;
QY 1 MSHFAVIADEPFSSHVRALONLAQELVARGHRTVFQQHDC--KALVTGSDIGFQTVGLQT 58
DB 58 MAHIAFILPAAGHVMTGLVGAELAAARHRTYALPEDMADRAVRVGA----- 106
QY 59 HPPGSLHLLHAAHPLGSMRLINEMARTSDMLCRE----- 96
DB 107 -----RAVTPLDREFR-----ADVPKESEYTDGEFLKVLWLLDTT 148
QY 97 ---LPAAFHALQIEGVIVDMPBA---GAVVAEASGLPFVSVACALPLNREPGLPLAVM 149
DB 149 ADTLPILLESAPABDRDDVANDPSTFTGLLLAGKWDIPVIRSTPVSANHNAH--P 205
QY 150 PFYGTSD---AARERTYTSKYLMLMRH---DRVIAHACRMG-----LAPR----- 193
DB 206 PFEEGAQVDPALIELTARAER---LLKEHGTSDPVAFAATVQSGPLFYNPYFOYA 261
QY 194 ----EKLHCFSPQLAQISQLISELPDFPKALPDCFHAVGLRQPGTSGSSTSYFRSPDK 249
DB 262 GETFDREHVEGCA-----PRAS-----PH-----GTWRPEDGR 292
QY 250 PRIFASLGLTQGHRYGLFTTIKACEEVDQAQLLAHOGGLSATQAGELARGDQIQQVDF 309
DB 293 PLVWSLGIYINERPGIFRACVAFDRPWNILLVGLGGLGAGDLGPLPE--NVLVRDFV 350
QY 310 DQSAALSQLQTLTHGGMTVLDASTRTPILALPLAFDQPGVASRIYTHGIGK-RASRF 368
DB 351 PLGDLVPLHTDLLVNGGTSTANEALAGVPIVAMPEPEPRATARRIAELDGLDWLLPGE 410
QY 369 TTSHALARQIRSLNTDYPQMTKIOAALRAGGTAAADIVE 412
DB 411 VTAEKLSGAQRVLTDDRIRKGLDRMRGEIRRAGGPAVAADVE 454

RESULT 10

Q81EB6
ID Q81EB6 PRELIMINARY; PRT; 402 AA.
AC Q81EB6;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.-).
GN BC2066
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=226900;
XN [1]
XN SEQUENCE FROM N.A.
XN MEDLINE=22608415; PubMed=12721630;
XN Ivanova N., Sorokin A., Anderson I., Galleron N., Gandelon B.,
XN Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
XN Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
XN Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
XN Overbeek R., Kyripides N.;
XN "Genome sequence of Bacillus cereus and comparative analysis with
XN Bacillus anthracis.";
XN Nature 423:87-91 (2003).
XN EMBL; AE017004; AAP09035.1;
XN GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
XN GO; GO:0008152; P:metabolism; IEA.
XN InterPro; IPR006326; UDPGT_MGT.
XN InterPro; IPR002213; UDP_Gluco_trans.
XN Pfam; PF00201; UDPGT; 1.
XN TIGRFAMs; TIGR01426; MGT; 1.
XN Glycosyltransferase; Transferase; Complete proteome.
XN SEQUENCE 402 AA; 45743 MW; 75B9D25EBC38D82 CRC64;
QY 1 MSFAVIADEPFSSHVRALONLAQELVARGHRTVFQQHDCALVTGSDIGFQTVGLQTHP 60
DB 1 MANVLVINFPGEGHINPTLAVVSELVQRGETVVSVCIEDYRKKEVATGAEFVFP----- 54
QY 61 PGSLHLLHAAHPLGSMRLINEMARTSDMLCRSLPAAFHALQIEGVIVDMPBAGAV 120
DB 55 ENFUSQINIMERVNEGSGPLTSLHMEASERIVTQIVETKEEKYDYLINHPVGR 114
QY 121 VAEASGLPFVSVACALPLNR-----EPCLPLAVMPPEYGTSDAARERTYS-----EK 168
DB 115 IVNLTQLPSVSSCTTFVAVNQYINPHDEQSRQVDEMDPLYQSLAGMERWNRKQGMKNS 174
QY 169 IYDMLMRHDEVTAAHACRMGLAPRKLHHCPSPLAQISQLIPELDPPKALPDCFHAVG 228
DB 175 MYDLMNHGPDITVI-----TSKE-----YQPRSEV-----FDESKYKVG 209
QY 229 PLRQPGTSGSSTSYFRSPD---KPRIFASLGLTQGHRYGLFTTIKACEEVDQAQLLAH 285
DB 210 PSIAIRKEVGS---PFTEDLKNKVIPIFISMGTVFNEQPALYKEKCFEAFQVDATVVLV 265
QY 286 CGGLSATQAGELARGDQIQQVDFADQSAALSQLITHTGKMTVLDASTRTPILALPL 345
DB 266 GKKINISQFENIK--NFKLYNTVPLEVLOHADVFTVTHGGMNSSEALYGVPLVPIV 323
QY 346 AFDQPGVASRIYTHGIGKRASR-FTTSHALARQIRSLNTDYPQMTKIOAALRAGGT 404
DB 324 TGDQPFVAKSLTEVGAGITLNRNELATSELLRETIVKVKMDVTFKENSILKVGESLRNAGY 383
QY 405 PAAAD 409
DB 384 QRAVE 388

RESULT 11

398 LRLAGTTPAAADIVEQAMRT 417
368 AQTAGNTLQAQYIINKDLQT 397

RESULT 13

933H13
ID Q93H13 PRELIMINARY; PRT; 379 AA.
AC Q93H13
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative glycosyl transferase.
GN Streptomyces avermitilis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21477403; PubMed=11572948;
RX Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
Shinose M., Takahashi Y., Horioka H., Nakazawa H., Osonoe T.,
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
"Genome sequence of an industrial microorganism Streptomyces
avermitilis: Deducing the ability of producing secondary
metabolites";
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMEL; AB070941; BAE69204.1; .
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006326; UDPGT_MGT.
DR Pfam; PF00201; UDPGT_1.
DR TIGRfams; TIGR01426; MGT; 1.
DR Transferase.
KW TRANSFERASE.
SQ SEQUENCE 379 AA; 38968 MW; B5PED865CB111DSC CRC64;

Query Match 9.8%; Score 219.5; DB 2; Length 379;
Best Local Similarity 26.5%; Pred. No. 2.6e-08;
Matches 109; Conservative 60; Mismatches 172; Indels 71; Gaps 18;
QY 22 AQLVARGHRTVFQHQDCKALVTGSDIGFQTVGLQTHPPGSLSHLLHLAHP;GPSMLR 81
Db 4 AALAAARGHDIAWGHPELVRLGLAGADAVVFPAL---PEDGLSRPAGLK---GPAAFQ 56
QY 82 LINE--MARTSDMLCRLPAFHALQIEGVIVDQMERAGAVVAESGLPFFSVACALPLN 139
Db 57 FLWSEFLAVLADAMPQVRAATEAYDPDVVCDQAVACALVAESLGRPWTSATT---S 113
QY 140 REPLGLAVWPPEYGTSDAAREYTTSEKIYDMLMRHDRVIAHACRM---GLA-PRE 194
Db 114 AELVDPLAMP-----KVAWL-----DGLGLBELRRITGGAGADPR- 151
QY 195 KLHCFSPALAQISOLIPELDFFRKALPCFHAVGP--LRPQGTGSGSTSYFFSPDKPRI 252
Db 152 -----PSPGVLAITYTTRALLGPVE-LPRVMLVGPSVAARPAGPDPFPEWLEASALPTV 205
QY 253 FASLGLTQCHRYGLRTIAKCEVDQAQLLHACGSL-ATQAGELARGGDIQ----- 304
Db 206 LWSLGTANDAGARFLNAR-----AGAL-----GGIADRVAVLVDPGGVWEHPVPTV 254
QY 305 -VYDFADQSAALSQAQLTTHGGMNTVLDAISKRTPLALPLAPDQGVASRIYTHGIGK 363
Db 255 LVRRYVLPQALLERLDVAVVCHAGINTVCEALWHGVLPVAPIRDDQIVAAQVVDAGAV 314
QY 364 RASRETTTSHA--LARGIRSLTNTDYPQRYTK-IQAALLRAGGTPAAADIVE 412
Db 315 RL-RFGRAAARIGAAVEAVLPAQCHRRKAAEAVGESFRAAGSGESAADRLIE 365

RESULT 14

933H13
ID Q93H13 PRELIMINARY; PRT; 417 AA.
AC Q93H13
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Macrolide glycosyl transferase.
GN MGTA.
OS Streptomyces ambofaciens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC23877;
RX MEDLINE=99445176; PubMed=10517588;
RA Pernodet J.L., Gourmelin A., Blondelet-Rouault M.H., Cundliffe E.;
"Dispensable ribosomal resistance to spiramycin conferred by srmA in
the spiramycin producer Streptomyces ambofaciens";
Microbiology 145:2355-2364(1999).
RL EMEL; AJ223970; CHA11707.1; .
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006326; UDPGT_MGT.
DR InterPro; IPR002213; UDP_gluco_trans.
DR TIGRfams; TIGR01426; MGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
DR Transferase.
KW TRANSFERASE.
SQ SEQUENCE 417 AA; 45427 MW; B693EF18F6FCASFD CRC64;

Query Match 9.8%; Score 219; DB 2; Length 417;
Best Local Similarity 24.8%; Pred. No. 3.2e-08;
Matches 111; Conservative 68; Mismatches 185; Indels 84; Gaps 17;
QY 2 SHPAVIAPPPFFSHVRLQNLQAQLVAEGRHVTF-----PQHDCKALVTGSDIGFQTVGL 56
Db 22 AHAIFPSIRAHGHVNPVSLVIRELVARGHRTVTAIDPLFAE---KVAETGAEPKLWNSTL 78
QY 57 QTHPPGSLSHLLHLAHP--LGPSMLRLINEMARTSDMLCRLPAFHALQIEGVIVDQM 114
Db 79 ----RGP-----DADPDWAGTTPLDNVBPFLDDAIQALPOLIAAYEGDEPDLVLDIT 127
QY 115 EPAGAVVAESGLPFFSVACAL---PLNREPLGLAVWPFF--EYGTSDAAREYTTSP- 167
Db 128 SYPARVLAAHGWGPAVSLSPNLVAMGYEEVGRPTWEEPLKTERGRAYDARFGMLKEN 187
QY 168 ---KIYDMLMRHDRV-----VIAHACRMGLAPREKLHCFSPALAQISOLIPELDPR 217
Db 188 GUTEDPDPFFVGRDPSLVLIPKALQPHADRV-----DEKTHTFVGACQGDRA----- 234
QY 218 KALPDCTFHAVGLPQPGTQSGSTSYFFSPDKPRIFASLGLTQCHRYGLRTIAKACEV 277
Db 235 -----ABGDWRRPESGA-----EKVVLVLSIGSSFTTRKPAFYACVFAFGAL 274
QY 278 DAQULLAHCG-GLSATQAGELARGGDIQVYDFADQSAALSQAQLTTHGGMNTVLDAIAS 336
Db 275 PGWHVVLQVGRHVDPAELGDVPE--NVEVRSWVVPOLAILKQADLFVTHAGAGSGQGLAT 332
QY 337 RPTLLALPLAFTDQGVASRIYHIGICKRASRTTSSALARQIR-----SLITWTDYPOMT 392
Db 333 ATPIVAVPQAVDQFGNADMLQGHGVGR---HLPTERTABALRAAGLALVEDPEFVARLK 389
QY 393 KIAALRAGGTTPAAADIVE---QAVRT 417
Db 390 EIQAGMAREGGTTRADLTAEALAAAT 417
RESULT 15
O68841
ID O68841 PRELIMINARY; PRT; 424 AA.
AC O68841
DT 01-AUG-1998 (TREMELrel. 07, Created)

Job time : 40.4397 secs

```
I 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
R Glycosyltransferase OleI.
N OLEI.
S Streptomyces antibioticus.
C Bacteria; Actinobacteridae; Actinobacteriales;
C Streptomycineae; Streptomycetaceae; Streptomycetes.
K NCBI_TaxID=1890;
N N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 11891;
X MEDLINE=98343801; PubMed=9880207;
A Quiros L.M., Aguirrezabalaga I., Olano C., Mendez C., Salas J.A.;
L Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
R EMBL; AF055579; AAC12648.1; -.
R PTR; T51110; T51110.
R GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0008152; P:metabolism; IEA.
R InterPro; IPR006326; UDPGT_MGT.
R InterPro; IPR002213; UDP_gluco_trans.
R Pfam; PF00201; UDPGT_1.
R TIGRFAMs; TIGR01426; MGT; 1.
R PROSITE; PS00375; UDPGT; 1.
W transferase.
Q SEQUENCE 424 AA; 45384 MW; FE7F90DB9A1C11EC CRC64;

Query Match 9.7%; Score 217; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 4.6e-08;
Matches 111; Conservative 69; Mismatches 185; Indels 84; Gaps 19;

Y 3 HFVIAAPPFPHVRLQNLQAEVLVARGHVTFFQHDCKALV-----TCSDIGFQTVGLQ 57
b 14 HIGFFNIPGHGVNPSLIGVQGLVARGQVSIGITDFGCAQVKGRAVAVYGF----- 67
Y 58 THPPGSLSHLHLAHPGLGPMRLINEMARTSDMLC---RELPAAFHALQTEG---VIV 111
b 68 -----ILPEEFNPEELIAEDQSRWACSLAEAFVLPQLRTATPTTGRDLIVY 115
Y 112 DMEPAGAVVAEASGLPFV-----SVACA-----LPLNEPGLPL---AVMPFEYGTSDA 158
b 116 DIASWPAPVLGRKWDIPFVQLSPFVSAYEGFEDVPAVQDPTADRGEEAAAP--AGTGDA 173
Y 159 ARERYTTSKIIYDMLRRHDRVIA---HH-----ACRMGLAPREKLHCFSPLAQISQLI 210
b 174 --BEGAEAE---DGLVRFTRLSAFLEEGVDTPATEFLIAPNRCIVGCRAP----SQIK 224
Y 211 PELDFPRKALPDCFHVAGPL---RQPQGT---PGSSTSYFPSPDKPRIFASLGTQGHRY 264
b 225 GD-----TVGDNVTFVGPTYGDRSHQGTWEGFGHG-----RVLIIALGSAFTDEL 270
Y 265 GLPFTIAKAEEDVAQLLLAHCGLSATQAGELARGDILQWDFADQSALSOAQLTIH 324
b 271 DFVYTCLSAVDGLDHWVLSVGRFVDPADLGEVP--PNVEVHOMVQDLILTRASAFITH 328
Y 325 GGNVTVLDAISRTPLALPLAFDPQGVASRIYVHGIGKRASR-FTTSHALARQIRSLIT 383
b 329 AGKGTWEALSNAPVPMVAVPQIAEOTMNAERIVELGLGRHIPRQDVTAETKLEAVLAVAS 388
Y 384 NTYPQRTKIQAALRLAGCTPAADIVE 412
b 389 DPGVAERLAARQEIIRAGGARAADILE 417
```

Search completed: February 29, 2004, 14:50:54

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

on on: February 29, 2004, 14:27:19 : Search time 7.26404 Seconds
(without alignments)
3089.496 Million cell updates/sec

file: US-09-941-947A-28
sequence: 1 MSHFAVIAPPPFHVRLQN.....EOMRTCPVLGQDYATATL 431

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
aximum DB seq length: 2000000000

set-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1790	80.2	431	1 CRTX PANAN	P21686 pantoea ana
2	1018.5	45.7	413	1 CRTX ERWHE	Q01330 ewinia her
3	228.5	10.2	430	1 OLSD STRAT	O53685 streptomyc
4	183	8.2	502	1 UDCL RABIT	P36514 cryptolagus
5	181.5	8.1	418	1 MGT STRLI	Q54387 streptomyc
6	174.5	7.8	541	1 CGT HUMAN	Q16880 homo sapien
7	167	7.5	455	1 UPOG HORVU	P14726 hordeum vul
8	164.5	7.4	541	1 CPT MOUSE	Q64676 mus musculu
9	164.5	7.4	541	1 CPT RAT	Q09426 rattus norv
10	162	7.3	535	1 UD11 RAT	Q64550 rattus norv
11	158	7.1	414	1 YF24 MYCTU	Q50583 mycobacteri
12	156	7.0	530	1 UD88 RAT	Q62789 rattus norv
13	154.5	6.9	527	1 UD11 HUMAN	Q944x1 homo sapien
14	150.5	6.7	505	1 UGTF CABEL	Q18081 caenorhabdi
15	150.5	6.7	530	1 UD19 HUMAN	O60656 homo sapien
16	150	6.7	528	1 UD81 NACFA	Q9xt55 macaca fasc
17	149	6.7	533	1 UD12 MOUSE	P70691 mus musculu
18	148.5	6.7	471	1 UPO2 MAIZE	P16165 zea mays (m
19	147.5	6.6	530	1 UD1A HUMAN	Q9haw8 homo sapien
20	145.5	6.5	471	1 UPO3 MAIZE	P16167 zea mays (m
21	145.5	6.5	530	1 UD12 HUMAN	P36509 homo sapien
22	143.5	6.4	471	1 UPO1 MAIZE	P16166 zea mays (m
23	142	6.4	529	1 UD8S HUMAN	Q9by64 homo sapien
24	140.5	6.3	530	1 UD17 HUMAN	Q9haw7 homo sapien
25	139.5	6.3	379	1 YG36 METWA	Q8pwf3 methanosarc
26	139.5	6.3	527	1 UDAL RAT	P36510 rattus norv
27	138.5	6.2	530	1 UD8F HUMAN	P54855 homo sapien
28	138	6.2	533	1 UD12 RAT	P20720 rattus norv
29	137	6.1	533	1 UD11 HUMAN	P22309 homo sapien
30	134	6.0	531	1 UD13 RAT	Q64637 rattus norv
31	134	6.0	531	1 UD15 RAT	Q64638 rattus norv
32	133.5	6.0	361	1 MURG CAUCR	Q8a5al caulobacter
33	133.5	6.0	388	1 Y453 METAC	Q8tt10 methanosarc

ALIGNMENTS

RESULT 1	CRTX PANAN	STANDARD;	PRT;	431 AA.
ID	CRTX PANAN			
AC	P21686;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DB	Zeaxanthin glucosyl transferase (EC 2.4.1.-).			
GN	CRTX OR UGT101.			
OS	Pantoea ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoea.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
RX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashina K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
CC	- FUNCTION: CATALYZES THE GLYCOSYLATION REACTION WHICH CONVERTS			
CC	ZEAXANTHIN TO ZEAXANTHIN-BETA-DIGLUCOSIDE.			
CC	- PATHWAY: Carotenoid biosynthesis.			
CC	- SIMILARITY: Belongs to the UDP-glycosyltransferase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D90087; BAAL4125.1; -			
DR	PIR; B37802; B37802.			
DR	InterPro; IPR002213; UDP_gluco_trans.			
DR	Pfam; PF00201; UDPGT; 1.			
DR	PROSITE; PS00375; UDPGT; 1.			
KW	Transferase; Carotenoid biosynthesis.			
SK	SEQUENCE 431 AA; 47241 MW; 9CE7A798608BC9D CRC64;			
	Query Match 80.2%; Score 1790; DB 1; Length 431;			
	Best Local Similarity 80.5%; Pred. No. 4.7e-131;			
	Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;			

QY	1	MSHFAVIAPPPFHVRLQNLAQELVAERHRTVFPQDCHCALVTGSDIGFTVGLQTHP	60
DB	1	MSHFAAIAFPFHVRLQNLAQELVAERHRTVFPQDCHCALVTGSDIGFTVGLQTHP	60
QY	61	PGSLSHLHLAHPGPSMLRLINEMATSNLRELPAAFHALQIEGVIVDQEPAGAV	120
DB	61	PGALTRVHLAHPGPSMLRLINEMATSNLRELPAAFHALQIEGVIVDQEPAGAL	120

Q9haw9 homo sapien
Q7vep8 mycobacteri
Q91280 pleuronecte
Q63886 mus musculu
O77649 macaca fasc
P36537 homo sapien
O19103 cryptolagus
O64634 rattus norv
O08224 mycobacteri
P36513 cryptolagus
Q8pzb2 methanosarc
P09875 rattus norv

[illegible]

```

Query Match      45.7%; Score 1018.5; DB 1; Length 413;
Best Local Similarity 48.8%; Pred. No. 2e-71;
Matches 211; Conservative 62; Mismatches 128; Indels 31; Gaps 5;

```

[illegible]

PROSITE; PS00375; UDPGT; 1.
Transferase; Glycosyltransferase; Antibiotic resistance.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

CC ENBL: L01083; AAA18023.1; -
DR InterPro: IPR002213; UDP_glucosyl_trans.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Multigene family; Microsome.
FT NON TER 1
FT TRANSMEM 466 481 POTENTIAL.
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 502 AA; 57449 MW; 56656705FAEID35 CRC64;

Query Match 8.2%; Score 183; DB 1; Length 502;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;
Matches 97; Conservative 75; Mismatches 177; Indels 98; Gaps 19;

Qy 12 FSHVRLONLAQELVARGHRTVFQQHDCALVTGSDIGPQTVGLQTHPE-----GSLSH 66
Db 6 FSHWNLKVLLESLQURGHEITVLPSP-SLLDHTYKIPENVEVLQVTKETLMEELNT 64
Qy 67 LLHLAHLPLGFSM-----LRLINEMARTSDMLCRELPAAAFHALQIE 107
Db 65 VLYMSSFEEL-PTLSWKKVLGKVMGKQFSKLLRRVCDSAITNKKELDLRLKAA----KFD 119
Qy 108 GVIVDQMEPAGAVVABASGLPFFV-----SVALPLNREPGCLPLAVMPFEV--G 154
Db 120 ICLADPLAFGCELVAELNIPFVYSFRFSIGNIERSCA-----GLP--TPSSVPG 169
Qy 155 TSDAARERYTTSEKIYDMLM-RRHDRVIAHH-----ACR-MGLAPRE 194
Db 170 STSGLTNNMNFVQELKNWLLYLNMDMFSHEMLSEWDEYYSKVLGRRTTICEIMGKAEW 229
Qy 195 KLHHCFSPLAQISOLIPELDPKALPDCHAVGPL--RQPGTSGSTSVFSP--DKP 250
Db 230 LIRSFW-----DFEPPFPFLFN-FEYVGLGHCKPAKPUFEELEBEVQSGNDGV 277
Qy 251 RIFASLGTQGHRYGLFRTIAKACEEVDQAOLLAHCGSLGATAGELARGDIOVDPAD 310
Db 278 VVFTLGSMIQNLTEERSNLASALQIPQKVLWRYTGKPAF-----LGNTRLFIEWIP 331
Qy 311 QSAALS--QMLTITHGGMNTVLDIASRTPLALPLAFOPGVASRIVYHIGIKRAS-R 367
Db 332 QNDLLGHCPKTRAFITHGNTGLYEALYHGVPMVGIPLFGDQPDNIARVAKAGAAVDVLR 391
Qy 368 FTTSHALARQIRSLTWTYDQPMTKI 394
Db 392 IMTSSLLKALKDWINNPSYKENDAKL 418

RESULT 5
MCT_STRLI STANDARD; PRT; 418 AA.
AC Q54387;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.1).
GN MGT OR UGT102A1.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RX MEDLINE=92104506; PubMed=1761231;
RA Jenkins G., Cundliffe B.;
RT "Cloning and characterization of two genes from Streptomyces lividans that confer inducible resistance to lincomycin and macrolide

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

CC ENBL: L01083; AAA18023.1; -
DR InterPro: IPR002213; UDP_glucosyl_trans.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Multigene family; Microsome.
FT NON TER 1
FT TRANSMEM 466 481 POTENTIAL.
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 502 AA; 57449 MW; 56656705FAEID35 CRC64;

Query Match 8.2%; Score 183; DB 1; Length 502;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;
Matches 97; Conservative 75; Mismatches 177; Indels 98; Gaps 19;

Qy 12 FSHVRLONLAQELVARGHRTVFQQHDCALVTGSDIGPQTVGLQTHPE-----GSLSH 66
Db 6 FSHWNLKVLLESLQURGHEITVLPSP-SLLDHTYKIPENVEVLQVTKETLMEELNT 64
Qy 67 LLHLAHLPLGFSM-----LRLINEMARTSDMLCRELPAAAFHALQIE 107
Db 65 VLYMSSFEEL-PTLSWKKVLGKVMGKQFSKLLRRVCDSAITNKKELDLRLKAA----KFD 119
Qy 108 GVIVDQMEPAGAVVABASGLPFFV-----SVALPLNREPGCLPLAVMPFEV--G 154
Db 120 ICLADPLAFGCELVAELNIPFVYSFRFSIGNIERSCA-----GLP--TPSSVPG 169
Qy 155 TSDAARERYTTSEKIYDMLM-RRHDRVIAHH-----ACR-MGLAPRE 194
Db 170 STSGLTNNMNFVQELKNWLLYLNMDMFSHEMLSEWDEYYSKVLGRRTTICEIMGKAEW 229
Qy 195 KLHHCFSPLAQISOLIPELDPKALPDCHAVGPL--RQPGTSGSTSVFSP--DKP 250
Db 230 LIRSFW-----DFEPPFPFLFN-FEYVGLGHCKPAKPUFEELEBEVQSGNDGV 277
Qy 251 RIFASLGTQGHRYGLFRTIAKACEEVDQAOLLAHCGSLGATAGELARGDIOVDPAD 310
Db 278 VVFTLGSMIQNLTEERSNLASALQIPQKVLWRYTGKPAF-----LGNTRLFIEWIP 331
Qy 311 QSAALS--QMLTITHGGMNTVLDIASRTPLALPLAFOPGVASRIVYHIGIKRAS-R 367
Db 332 QNDLLGHCPKTRAFITHGNTGLYEALYHGVPMVGIPLFGDQPDNIARVAKAGAAVDVLR 391
Qy 368 FTTSHALARQIRSLTWTYDQPMTKI 394
Db 392 IMTSSLLKALKDWINNPSYKENDAKL 418

RESULT 5
MCT_STRLI STANDARD; PRT; 418 AA.
AC Q54387;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.1).
GN MGT OR UGT102A1.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RX MEDLINE=92104506; PubMed=1761231;
RA Jenkins G., Cundliffe B.;
RT "Cloning and characterization of two genes from Streptomyces lividans that confer inducible resistance to lincomycin and macrolide

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

CC ENBL: L01083; AAA18023.1; -
DR InterPro: IPR002213; UDP_glucosyl_trans.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Multigene family; Microsome.
FT NON TER 1
FT TRANSMEM 466 481 POTENTIAL.
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 502 AA; 57449 MW; 56656705FAEID35 CRC64;

Query Match 8.2%; Score 183; DB 1; Length 502;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;
Matches 97; Conservative 75; Mismatches 177; Indels 98; Gaps 19;

Qy 12 FSHVRLONLAQELVARGHRTVFQQHDCALVTGSDIGPQTVGLQTHPE-----GSLSH 66
Db 6 FSHWNLKVLLESLQURGHEITVLPSP-SLLDHTYKIPENVEVLQVTKETLMEELNT 64
Qy 67 LLHLAHLPLGFSM-----LRLINEMARTSDMLCRELPAAAFHALQIE 107
Db 65 VLYMSSFEEL-PTLSWKKVLGKVMGKQFSKLLRRVCDSAITNKKELDLRLKAA----KFD 119
Qy 108 GVIVDQMEPAGAVVABASGLPFFV-----SVALPLNREPGCLPLAVMPFEV--G 154
Db 120 ICLADPLAFGCELVAELNIPFVYSFRFSIGNIERSCA-----GLP--TPSSVPG 169
Qy 155 TSDAARERYTTSEKIYDMLM-RRHDRVIAHH-----ACR-MGLAPRE 194
Db 170 STSGLTNNMNFVQELKNWLLYLNMDMFSHEMLSEWDEYYSKVLGRRTTICEIMGKAEW 229
Qy 195 KLHHCFSPLAQISOLIPELDPKALPDCHAVGPL--RQPGTSGSTSVFSP--DKP 250
Db 230 LIRSFW-----DFEPPFPFLFN-FEYVGLGHCKPAKPUFEELEBEVQSGNDGV 277
Qy 251 RIFASLGTQGHRYGLFRTIAKACEEVDQAOLLAHCGSLGATAGELARGDIOVDPAD 310
Db 278 VVFTLGSMIQNLTEERSNLASALQIPQKVLWRYTGKPAF-----LGNTRLFIEWIP 331
Qy 311 QSAALS--QMLTITHGGMNTVLDIASRTPLALPLAFOPGVASRIVYHIGIKRAS-R 367
Db 332 QNDLLGHCPKTRAFITHGNTGLYEALYHGVPMVGIPLFGDQPDNIARVAKAGAAVDVLR 391
Qy 368 FTTSHALARQIRSLTWTYDQPMTKI 394
Db 392 IMTSSLLKALKDWINNPSYKENDAKL 418

RESULT 5
MCT_STRLI STANDARD; PRT; 418 AA.
AC Q54387;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.1).
GN MGT OR UGT102A1.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RX MEDLINE=92104506; PubMed=1761231;
RA Jenkins G., Cundliffe B.;
RT "Cloning and characterization of two genes from Streptomyces lividans that confer inducible resistance to lincomycin and macrolide

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See

```

ZT      antibiotics, ";
XL      Gene 108:55-62(1991).
--      FUNCTION: SPECIFICALLY INACTIVATES MACROLIDES VIA 2'-O-
CC      GLYCOSYLATION USING UDP-GLUCOSE.
CC      -1- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
JR      EMBL; M74717; AAA26780.1; --
JR      PIR; JS0636;
JR      InterPro; IPR002213; UDP_gluco_trans.
JR      InterPro; IPR006326; UDPGT_Mgt.
JR      Pfam; PF00201; UDPGT; 1.
JR      TIGRFAMs; TIGR01426; MGT; 1.
JR      PROSITE; PS00375; UDPGT; 1.
KW      Transferase; Glycosyltransferase; Antibiotic resistance.
SQ      SEQUENCE 418 AA; 45682 MW; F7349765CEB8D2E4 CRC64;
-----
Query Match          8.1%; Score 181.5; DB 1; Length 418;
Best Local Similarity 22.7%; Pred.No.1.1e-06;
Matches 101; Conservative 67; Mismatches 186; Indels 91; Gaps 17;

ZY      2 SHFVIAFPFFSHVRALQNLQAELVARGHRVTFFQCHDKALVTGSDIGFTVGLQTHTPP 61
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      23 AHIAFMFSIALGHVNPSPVIRELSLVARGHRVTVAIPRLADKV--ABAGAEPKLMNSTLP 80
ZY      62 GSLSHELLHAAHP--LGPSMLRLINEMARTSDMLCRELPAAFHAIQEGVIVDOMEPAGA 119
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      81 GP-----DADPMAWGSTILDNVEFFLADAIQSLPLAQAYEGDEPDVLVHDIASYTA 133
ZY      120 VVAEASGLPFVSVA-CALP---LNREPGFLAMPFEYGTSDAARRRYTSKIIDWLMR 175
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      134 VLGRWRPEVPVISPCVMWAWEQEYGVGPMPWEPRKTERQAYANF-----HAWL-- 185
ZY      176 RHDRVIAHACRMGLAPREKLHCPSLAQISQLPELDPPRKALPDFHAVGPLPQPQG 235
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      186 BENGITDHDPPIGTPDRSIV-----LIP-----KAL-----QPHA 216
ZY      236 TFGSGTSYF-----PSPDKPRIFASLTGTLOGHRHYGLFRTIAKACEYD 278
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      217 DRVDDETITVFAGCQDQRTAEGDWARPEGAKEVWLVSLSGSTFKQPAFYRECVRATGELP 276
ZY      279 AQLLLAHCGGLSATQAG---ELARGD----IQVYDRADQSAALSQAQLTIHGGMNTVL 331
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      277 -----GMHTVLQGRHVDPALGDVPDNVEVTVWPOLAIQQADLFVTHAGAGGSQ 328
ZY      332 DALASRPTLIALLPADPQCPVASRIYVHGIGKRASFTTSHALARQIR-----SLLTNTDY 387
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      329 EGLATATPMIAVPAQADQFNADML--QGLGV-AETLPTTEATAKALTAAALAVDDPEV 385
ZY      388 PQRMTKIQAALRLAGGTPMAADIVE 412
DB       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      386 AARLKEIQARMAQEAQGTGPAULIE 410

RESULT 6
CGT_HUMAN ID CGT_HUMAN STANDARD; PRT; 541 AA.
AC Q16880; O00196;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-hydroxyacylphingosine 1-beta-galactosyltransferase precursor
DE (EC 2.4.1.45) (UDP-galactose-ceramide galactosyltransferase) {Ceramide
DE UDP-galactosyltransferase}
GN UGT8 OR CGT OR UGT4.
OS Homo sapiens (Human).
```

[illegible]

61 -----PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRSLPAAPH 102
77 FNSITSDAFASQKMNIFSGRLTAI-----ELFDILDHYTKCDLM-----VGNH 121
101 ALQIEG-----VIVDMPAGAVVAERAGLPFVSACALPLNRPFG--LPLAVMPFE 152
122 AL-IOGLKKEFDLLVDPNMOCGFVIAELLGKXVAFSTGLWYPAEYGAAPLAYVP-E 179
153 YQTSDAARERTTSKIDYMLMRHRDVRHIAHCRMGLA-----PREKLHCFSPLAQISQ 208
180 FNSLLTDRMN-----LIQRMXNGVYL1SRLGVSFLVLPK-----YERIMQKYN 223
209 LIPE-----LDPRKALPDCFHAGVGLRPOQGTGSGSTSYFSPD 248
224 LPEKSMVDLVHSGSLWMLCTDVALEFRPPLPNVYVGGILTKPAS-----PLPE 274
249 KPRI-PASLSTGLQGHRYGLFRTIAKACEVDQAQLLHACGGLSATQAGELAR----- 299
275 DIQRWNGANERHGFVLVSFGAGKVLSEDLANKL-----AGALGRLPKQVIRWF 323
300 -----GGDIQVDFADQSAALSOAQLT--ITHGGMNTVLDALASRTPLALLPLAFDQ 349
324 SQKPKNLGNNTKLIEWLPONDLLGHSKIKAPLSHGGLSIFETWYHGVVPGVPLFGDH 383
350 PGVASRIVYHIGKRAAS-RFTTSHALARQIRSLTNTDYPQRMTKIOAALRAGTTP 405
384 YQTRVQAKGMLLEWTKVTEKELYALVKVINNPYRQRAQKLSIHKDQCHP 440

RESULT 7

FOG HORVU STANDARD; PRT; 455 AA.
214726;
01-APR-1990 (Rel. 14, Created).
01-NOV-1997 (Rel. 35, Last sequence update).
28-FEB-2003 (Rel. 41, Last annotation update).
Flavonol 3-O-glucosyltransferase (SC 2.4.1.91) (UDP-glucose flavonoid
3-O-glucosyltransferase) (Bronze-1).
B21 OR UGT71A2.
Hordeum vulgare (Barley).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
C Triticeae; Hordeum.
X NCBI_TaxID=4513;
[1]
SEQUENCE FROM N.A.
P STRAIN=cy. Abyssinian 2231;
X MEDLINE=91329682; PubMed=2151660;
A Wise R.P., Rohde W., Salamini F.;
T "Nucleotide sequence of the Bronze-1 homologous gene from Hordeum
vulgare.";
L Plant Mol. Biol. 14:277-279 (1990).
C -1- FUNCTION: In the presence of other necessary color factors, this
C glycosylation reaction allows the accumulation of anthocyanin
C pigments.
C -1- CATALYTIC ACTIVITY: UDP-glucose + a flavonol = UDP + flavonol
C 3-O-D-glucoside.
C -1- PATHWAY: Anthocyanin biosynthesis.
C -1- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
C or send an email to license@sib-sib.ch).
C
C EMBL: X15694; CAA33729.1; --
R PIR: S14919; XUBHFC.
R InterPro: IPR002213; UDP_glucosyltrans.
R Pfam: PF00201; UDPGT; 1.
R PROSITE: PS00375; UDPGT; 1.

KW Transferase: Glycosyltransferase.
SQ SEQUENCE 455 AA; 47079 MW; PFD8BC2E2F103AC9C CRC64;
Query Match 7.5%; Score 167; DB 1; Length 455;
Best Local Similarity 24.1%; Pred. No. 1.6e-05;
Matches 112; Conservative 56; Mismatches 194; Indels 102; Gaps 21;
QY 3 HEAVLAPFPFFSHVRALQNLQAQLVLA---RGRVITFFQOQHD-----CKALVTGSDIGPQTV 54
DB 7 HIAVAVFPFSSEAVALFSAFALAAAPAGTSLSLFTTADNAAQLERKAGALPGNLRFFVEV 66
QY 55 --GLOTHPPGSLSHL-----LHLAHPGLSPMLRLINEMARTSDMLCRSLPAAFHALQ 105
DB 67 PDGV---PPGETSCLSPPRRMDLFMAAAEAGGVRLG-----EAAAC-----ASAGGAR 111
QY 106 IEGVIVDQMEPAGAVVAEASGLPFVSV----ACALPLNREPGLFLAVMPFFYGTSDAAAE 161
DB 112 VSCVVDGAPVWT-ADAAASAGAPWAVVTAASCAL-----LAHLRTDLAR- 155
QY 162 RYTTSEKIYDMLMRHRDVRHIAHCRMGLAPE-----KLHCFSPLA--QISQ-- 209
DB 156 ----RDVGDQASRADELLVAHAGLGGYVRDLDPGVVSGDFNVVISLLVHROAQLPK 210
QY 210 -----IPELDPP-----RKALPDCFHAGVGLR-POQGTGSGSTSYFSPS----- 246
DB 211 AATAVALNTFPGDLPDLIALAELPNCL-PLGYPHLLPGAEPTADTNEAPADPHGCLA 269
QY 247 -----PDKEPRI-PASLSTGLQGHRYGLFRTIAKACEVDQAQLLHACGGLSATQAGELARGG 301
DB 270 WLDRRPARSVAVVSGTNTATARPDELQELAAGLEASGAFFLMSLGGVVAAPRGFLERAP 329
QY 302 DIQVDFADQSAALSOAQLT--TITHGGMNTVLDALASRTPLALLPLAFDQPGVASRIV-Y 358
DB 330 GL-WVPWAPQVGLVRRHAAVGVAFVTRAGVASVMEGVSSGVPMACRFFPGDQTNARSVASV 388
QY 359 HGIGKRASTFTSHALARQIRSLTNTDYPQRMTKIOAALRLAG 402
DB 389 WGTSTAFDPMTRGAVANAVATLLRGEDGERMKAKAQELQAMVG 432
RESULT 8
CGT_MOUSE STANDARD; PRT; 541 AA.
AC Q64576; Q61634;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-hydroxyacylsphingosine 1-beta-galactosyltransferase precursor
DE (SC 2.4.1.45) (UDP-galactose-ceramide galactosyltransferase) (Ceramide
DE UDP-galactosyltransferase) (Cerebroside synthase).
GN UGT8 OR CGT OR UGT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c;
RA Bosio A., Binczek E., Stoffel W.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Coetzee T., Li X., Fujita N., Marcus J., Suzuki K., Francke U.,
RA Popko B.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE TRANSFER OF GALACTOSE TO CERAMIDE, A KEY
CC ENZYMIC STEP IN THE BIOSYNTHESIS OF GALACTOCEREBROSIDES, WHICH
CC ARE ABUNDANT SPHINGOLIPIDS OF THE MYELIN MEMBRANE OF THE CENTRAL
CC NERVOUS SYSTEM AND PERIPHERAL NERVOUS SYSTEM.
CC -1- CATALYTIC ACTIVITY: UDP-galactose + 2-(2-hydroxyacyl)sphingosine =
CC UDP + 1-(beta-D-galactosyl)-2-(2-hydroxyacyl)sphingosine.
CC -1- PATHWAY: Galactocerebroside biosynthesis.
CC -1- SIMILARITY: Belongs to the UDP-glycosyltransferase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X92122; CAAG3090.1; -
EMBL; X92123; CAAG3091.1; -
EMBL; X92124; CAAG3091.1; JOINED.
EMBL; X92125; CAAG3091.1; JOINED.
EMBL; X92126; CAAG3091.1; JOINED.
EMBL; X92177; CAAG3091.1; JOINED.
EMBL; U48896; AAC33576.1; -
EMBL; U48892; AAC33576.1; JOINED.
EMBL; U48893; AAC33576.1; JOINED.
EMBL; U48894; AAC33576.1; JOINED.
MGD; MGI:109522; Ugt8.
InterPro; IPR002213; UDP_gluco_trans.
Pfam; PF00201; UDPGT; 1.
PROSITE; PS00375; UDPGT; 1.
Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
Microsome.
SIGNAL 1 20
CHAIN 21 541
POTENTIAL
2-HYDROXYACETYLSPHINGOSINE 1-BETA-
GALACTOSYLTRANSFERASE.
TRANSMEM
472 492
CARBOHYD 78 78
CARBOHYD 333 333
CARBOHYD 442 442
CONFLICT 335 335
CONFLICT 458 458
CONFLICT 511 511
CONFLICT 524 524
CONFLICT 533 534
CONFLICT 541 541
SEQUENCE 541 AA; 61137 MW; B76F80A9B5326E8 CRC64;
Query Match 7.4%; Score 164.5; DB 1; Length 541;
Best Local Similarity 20.3%; Pred. No. 3.1e-05;
Matches 93; Conservative 76; Mismatches 197; Indels 85; Gaps 17;
QY 6 VIAPP--FSSHVRAQLAEVLVARGHRTVFPQHDCAVLTGSDIG--PQTVGLQTHP-- 60
D 24 IIVPEPMFSSHLVYFKTLASALHERGH-----HTVLLSEGRDIAPSNFYSLQRYPGI 76
QY 61 -----PGSLSHLLAAHPGLGSMRLINEMARTSDMLC--RELPA 100
D 77 FNSTTSDAELQSMENIFSGRLTAV-----ELVDILHYTKNCDMMVGNQALIQ 126
QY 101 FHAIQTEGVVDQMEPAGAVABASGLPVSACALPLNREP--LPLAVMPPEYGTSDA 158
D 127 LKSEKFDLLVDPNOMCGFVIAHLGVKYAVFTSLMYPAEVGPAPLAYVP--EFNSLLT 185
QY 159 ARERYTTSK-----IYDMERHRDVAHAHACRMGLAPREKLHCP--SPLAQ 205
D 186 DRNFMFLRMKNTGVYLLSRIQSVFLVLPKYERIMQ-----KYNLLPAKSMYDLVHGSLLWM 241
QY 206 ISQLIPDLPKALPDCHHAGVPLRQCGTSGSTSYFFSPDPKRPASLGLTQGHRYG 265
D 242 LCTDV-ALSPFPRLPLNVVYVGGILTKPAS-----PLPEDLRWVSGAQEHGFLV 291
QY 266 LPTIAKACEEVDQAALLAHCCGL-----SATOAGELARGGDIQVDFADQSAALSQ 317
D 292 SEGAGVYLYSEDIANKLAGALGRLPQKVIWFSQTKPKNL--GNNTKLIBELPQNDLLGH 349
QY 318 AQLT--ITHGGMNTVLDIASTPLALPLAFDPQGVASRVIVHGIGKGRASRTT--TSHAL 374
D 350 SNIRAPLSGGLNSIFETMVGHPVVGVPFLFGDHYDTMTVRQAKGMGILLENMTVTGEG 409
QY 375 ARQIRSLNTDYPQRMWTKIQALRLAGTGP 405

Db 410 YDALVYVNNPSYQRAQKLSEHKQPGHP 440
CGT RAT
ID CGT RAT STANDARD; PRT; 541 AA.
AC Q09426;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-HYDROXYACETYLSPHINGOSINE 1-BETA-GALACTOSYLTRANSFERASE precursor
DE (EC 2.4.1.45) (UDP-galactose-ceramide galactosyltransferase) (Ceramide
DE UDP-galactosyltransferase) (Cerebroside synthase).
OS UGT8 OR CGT OR UGT4.
GN Rattus norvegicus (Rat).
OC Sukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WLStar; TISSUE=Brain;
RX MEDLINE=94052143; PubMed=7694285;
RT "Ceramide UDPgalactosyltransferase from myelinating rat brain:
RT purification, cloning, and expression";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10265-10269 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=94358923; PubMed=7521399;
SA Stahl N., Jurevics H., Morel P., Suzuki K., Popko B.;
RT "Isolation, characterization, and expression of cDNA clones that
RT encode rat UDP-galactose: ceramide galactosyltransferase";
RL J. Neurosci. Res. 38:234-242 (1994)
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GALACTOSE TO CERAMIDE, A KEY
CC ENZYMAIC STEP IN THE BIOSYNTHESIS OF THE MYELIN MEMBRANE OF THE CENTRAL
CC ARE ABUNDANT SPHINGOLIPIDS OF THE MYELIN MEMBRANE OF THE CENTRAL
CC NERVOUS SYSTEM AND PERIPHERAL NERVOUS SYSTEM.
CC -!- CATALYTIC ACTIVITY: UDP-galactose + 2-(2-hydroxyacyl) sphingosine =
CC UDP + 1-(beta-D-galactosyl)-2-(2-hydroxyacyl) sphingosine.
CC -!- PATHWAY: Galactocerebroside biosynthesis.
CC -!- TISSUE SPECIFICITY: BRAIN, RESTRICTED TO THE OLIGODENDROCYTE-
CC CONTAINING CELL LAYERS OF CEREBRUM AND CEREBELLUM.
CC -!- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

Query Match 7.4%; Score 164.5; DB 1; Length 541;
Best Local Similarity 20.3%; Pred. No. 3.1e-05;


```

01-NOV-1997 (Rel. 35, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Pucative UDP-glucuronosyltransferase UGT1's precursor (EC 2.4.1.17)
(UNPGT).
UGT15 OR C18C4.3
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
UN SEQUENCE FROM N.A.
IC STRAIN-Bristol N2;
QA Gattung S.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[2]
UN REVISIONS.
RP Waterston R.;
IC Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL -! CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
XC beta-D-glucuronoside.
XC -! SIMILARITY: Belongs to the UDP-glucosyltransferase family.
XC
XC This SWISS-PROT entry is copyright. It is produced through a collaboration
XC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
XC the European Bioinformatics Institute. There are no restrictions on its
XC use by non-profit institutions as long as its content is in no way
XC modified and this statement is not removed. Usage by and for commercial
XC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
XC or send an email to license@isb-sib.ch).
XC
XC EMBL; U55369; AAK52183.1; -.
DR WormPep; C18C4.3; CB27363.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UNPGT; 1.
DR PROSITE; PS00375; UNPGT; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase; Glycoprotein;
KW Transmembrane; Signal; Multigene family.
KW SIGNAL 1
FT ?
FT CHAIN ? 505
FT PUTATIVE UDP-GLUCURONOSYLTRANSFERASE
FT UGT15.
FT
FT TRANSMEM 468 488
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 505 AA; 57244 MW; 7AA69450EFCDBB3B CRC64;
SQ
Query Match 6.7%; Score 150.5; DB 1; Length 505;
Best Local Similarity 20.4%; Pred. No. 0.00035;
Matches 91; Conservative 74; Mismatches 197; Indels 85; Gaps 19;
QY 13 SHVRLALQNLAEIARGHVRVFFQOH-----DCKAL----VTGSDIGFTVGLQTH 59
DB 9 SHVIQSQRIADELANAGHEVNVFPPDLNLTDKVPCKCKRRWPVTGLN-NYKFKKIQMG 67
QY 60 PGGSL-----SHLLHLAAHPLGFSMLRLINEMARTSDMLCELPAAAFALQISGIVD 112
DB 68 LSGDVFQSSQIWSKIFNTDSDPYQDEVTNCEEMVTNKEIKELKEKEDKA-----YIGE 122
QY 113 QMEPGAGVVAEASGLP--FVSVAICALPLNREPGL----PLAVMPFPGYTSDAARERYTTS 166
DB 123 QHLCHLGLAHILGIKHFMTIASCTMSVSRDSDIGITPSSLIFF-MSTLDATPAFFWQR 181
QY 167 EKIYDWLMRR-----HDRVIAAHACRMGLAPRKLIHCFSPLAQISQLI-----PELDFP 216
DB 182 AKNFVFLQMAHIRDEYRDVVLVNDMFKNFGSD---PFCVEFLAKTSDLIFFVSTDELEIQ 238
QY 217 RXALPDCHFAVGLRQPGTGGTSSTVSFPSPDKPR--IPASLCTLQ----- 261
DB 239 APTLGNVHV-IGGLGLSGGGGLDEKFPVKLMEKGVILFSLGTIANTNLPTIMENLM 297
QY 262 ---HRYGLFRTTAKACEVDAQLLHAHCGGLSAGQAGELARG-GDIQWVDFADQSAALISQ 317

```

298 KITQKFDYEFIIKV-DKFD-----RRSPDLAELGSLNVLVDMVDPQTAVLAH 343

318 AQLT---ITHGQNTVLDIASFTPLALPLAPDQPGVASRIYVHGIGKASEF---TTSH 372

344 PRLKAFITHAGYNSLWESAYAGVPVILIFFMFDQPRNGRSVERKGWILDRFLQIKDPD 403

373 ALARGISLLTNTDYPQRMKIQAAALR 399

404 ATEGAKEILVNPTQEKANRLKMLR 430

RESULT 15

UD19_HUMAN STANDARD; PRT; 530 AA.

AC O60656; Q9HAX0;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE UDP-glucuronosyltransferase 1-9 precursor, microsomal (EC 2.4.1.17)

DE UDP-glucuronosyltransferase 1a9 (UDPGT) (UGT1-9) (UGT1.9)

DE (UGT-11) (UGT11).

DE Homo sapiens (Human).

GN UGT1A9 OR UGT1 OR GNT1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RI [1]

RN SEQUENCE FROM N.A.

RP TISSUE=Liver;

RC Clotti M, Potter C., Owens I.S.;

RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RL [2]

RN SEQUENCE FROM N.A.

RM MEDLINE=21327373; PubMed=11434514;

RA Gong Q.H., Cho J.W., Huang T., Potter C., Gholami N., Basu N.K., Kunota S., Carvalho S., Pennington M.W., Owens I.S., Popescu N.C.;

RT "Three UDP-glucuronosyltransferase genes are encoded at the human UGT1 gene complex locus.";

RT Pharmacogenetics 11:357-368(2001).

RL [3]

RN SEQUENCE OF 1-285 FROM N.A.

RP Owens I.S., Gong Q., Cho J.W., Potter C., Gholami N.;

RA "Human phenol UDP-glucuronosyltransferase (UGT1A9) gene isozyeme exon 1.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases

CC -!- FUNCTION: UDPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds. This isoform has specificity for phenols.

CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor-beta-D-glucuronoside.

CC -!- SUBCELLULAR LOCATION: Microsomal.

CC -!- ALTERNATIVE PRODUCTS:

CC Even=Alternative splicing; Named isoforms=1;

CC Comment=A number of isoforms may be produced. Isoforms have a different N-terminal domain and a common C-terminal domain of 245 residues;

CC Name=1;

CC IsoId=O60656-1; Sequence=Displayed;

CC -!- TISSUE SPECIFICITY: Liver.

CC -!- SIMILARITY: Belongs to the UDP-glycosyltransferase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AF056188; AAC31425.1; -

DR EMBL; AF297093; AAG30418.1; -

DR EMBL; AF297091; AAG29816.1; -

DR Genew; HGNC:12541; UGT1A9.

```

MIM; 606434; -
MIM; 191740; -
GO; GO:0008152; p:metabolism; TAS.
InterPro; IPR002213; UDP_gluco_trans.
Pfam; PF00201; UDPGT; 1.
PROSITE; PS00375; UDPGT; 1.
Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
Multigene family; Microsome; Alternative splicing.
SIGNAL 1 25 POTENTIAL.
CHAIN 26 530 UDP-GLUCURONOSYLTRANSFERASE 1-9.
TRANSMEM 488 504 POTENTIAL.
CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 530 AA; 59940 MW; C417B9B86B403078 CRC64;

Query Match 6.7%; Score 150.5; DB 1; Length 530;
Best Local Similarity 19.4%; Pred. No. 0.00037;
Matches 91; Conservative 81; Mismatches 163; Indels 135; Gaps 20;

6 VIAPPPFSHVRLANLQELVARGHRTFFQCHDCKALVTGSDIGFO-----TVGLQTHP 60
29 LVPMDGSHFTGRSVVEKILRGHEV-----VVVMPVSNQLSNCTVKTY- 77
61 PGSLSHLH-----LAAHPLGPSMLRLINEMARTS-----DML-----CREL----- 97
78 --SYTYLDDLRBFKAFHQAQVRSIYSLMGSYNDIFDLFFSNCSLFPKDKLVE 135
98 ---PAAFALQIRGVVDQMPAGAVVAASGLPFSVA-----CALPLNRE 141
136 YLKESFDA-----VFLPPDNCGLIVAKYFSLPSVVFARGILCHYLEGAQCPAPLSYV 190
142 PGLPLAVMPPEYGTSDARERTTSEKIYDMLRRHRDRIAHACRMGLAPREKLHCFS 201
191 PRILL-----GFSDA-----MTFKEVRNHHLEHLLCHRRFFKNALIASEILQ--T 237
202 PLAQ-----ISQLPELPPRKALPDCFHAVGELRQPOQTP----- 237
238 PTEYDLKSHTSIMLRDFTDVLDPKVPWPNMIP-IGSINCHQKPLPMBFEAYINASGE 296
238 -----GSSTSYFPPSPDKPRIFASLGTLOCHRYGLFRTIAKACEVDACLHACGGL 289
297 HGIWVFLSGSMVSEIPEKKAMAIADALG-----KIPQTVLWRYTG-- 336
290 SATQAGELARGDIOVDFADQSAALS--QAQLTIHGGMMTVLDAIASRTPLIALPLAF 347
337 --TRPSNLA--NNTILVKWLQNDLLGHPMTRAFITHAGSGVYVESICNGVPMWMLFG 392
348 DQGVASRIVVHGIGKASRP-TTSHALARQIRSLLTNTDYPQMTKIQ 396
393 DQMDNAKRMETKAGCVTLNVLEMTSEDLNALKAVINDKSYKENIMRLSS 442
```

sarch completed: February 29, 2004, 14:45:06
DB time : 10.264 secs

GerCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

February 29, 2004, 14:51:24 ; Search time 28.6159 Seconds
(without alignments)
3180.293 Million cell updates/s

US-09-941-947A-28

```

FIECC score: 2231
1 MSHFAVIAPPPFSHVRLQN.....EQAMRTCQPVLSGQDYATAL 431
sequence:

```

spring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

809742 segs, 211153259 residues

total number of hits satisfying chosen parameters: 809742

Maximum DB seq length: 0

```

Maximum DB seq length: 0
Maximum DB seq length: 2000000000

```

st-processing: Minimum Match 0%

et-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

awase :	
Furukawa	6/pctodata/2/pubppaa/US09A_PUBCOMB.pcp.*
1:	/cgm2_6/pctodata/2/pubppaa/US07_PUBCOMB.pcp.*
2:	/cgm2_5/pctodata/2/pubppaa/PCT_NEW_PUB.pcp.*
3:	/cgm2_6/pctodata/2/pubppaa/US06_NEW_PUB.pcp.*
4:	/cgm2_6/pctodata/2/pubppaa/US06_PUBCOMB.pcp.*
5:	/cgm2_6/pctodata/2/pubppaa/US07_NEW_PUB.pcp.*
6:	/cgm2_6/pctodata/2/pubppaa/PCTUS_PUBCOMB.pcp.*
7:	/cgm2_6/pctodata/2/pubppaa/US08_NEW_PUB.pcp.*
8:	/cgm2_6/pctodata/2/pubppaa/US08_PUBCOMB.pcp.*
9:	/cgm2_6/pctodata/2/pubppaa/US09A_PUBCOMB.pcp.*
10:	/cgm2_6/pctodata/2/pubppaa/US09B_PUBCOMB.pcp.
11:	/cgm2_6/pctodata/2/pubppaa/US09C_PUBCOMB.pcp.
12:	/cgm2_6/pctodata/2/pubppaa/US09_NEW_PUB.pcp.*
13:	/cgm2_6/pctodata/2/pubppaa/US10A_PUBCOMB.pcp.
14:	/cgm2_6/pctodata/2/pubppaa/US10B_PUBCOMB.pcp.
15:	/cgm2_6/pctodata/2/pubppaa/US10C_PUBCOMB.pcp.
16:	/cgm2_6/pctodata/2/pubppaa/US10_NEW_PUB.pcp.*
17:	/cgm2_6/pctodata/2/pubppaa/US60_NEW_PUB.pcp.*
18:	/cgm2_6/pctodata/2/pubppaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2231	100.0	431	10	US-09-941-9478-28		Sequence 28, Appl
2	2231	100.0	431	14	US-10-218-119-4	2	Sequence 4, Appl
3	251.5	11.3	397	14	US-10-156-761-14886		Sequence 14886, A
4	247	11.1	412	14	US-10-156-761-8483		Sequence 8483, Ap
5	177	7.19	477	9	US-09-740-083-4		Sequence 3, Appl
6	167	7.5	527	9	US-09-562-678-2		Sequence 2, Appl
7	167	7.5	527	9	US-09-981-353-166		Sequence 166, App
8	167	7.5	527	13	US-10-052-586-522		Sequence 522, App
9	167	7.5	527	14	US-10-174-590-522		Sequence 522, App
10	167	7.5	527	14	US-10-156-758-522		Sequence 522, App
11	167	7.5	527	14	US-10-175-737-522		Sequence 522, App
12	167	7.5	527	14	US-10-173-706-522		Sequence 522, App
13	167	7.5	527	14	US-10-175-738-522		Sequence 522, App
14	167	7.5	527	14	US-10-175-452-522		Sequence 522, App
15	167	7.5	527	14	US-10-176-482-522		Sequence 522, App

ALIGNMENT'S

RESULT 1

```

US-09-941-947A-28
/ Sequence 28, Application US/09941947A
/ Publication No. US20030003528A1
/ GENERAL INFORMATION:
/ APPLICANT: Brzostowicz, Patricia C.
/ APPLICANT: Cheng, Qiong
/ APPLICANT: DiCossimo, Deana J.
/ APPLICANT: Koffas, Mattheos
/ APPLICANT: Miller, Edward S. Jr.
/ APPLICANT: Odom, J. Martin
/ APPLICANT: Picataggio, Steve
/ APPLICANT: Rouviere, Pierre E.
/ TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
/ FILE REFERENCE: CU1903 US NA
/ CURRENT APPLICATION NUMBER: US/09/941.947A
/ CURRENT FILING DATE: 2001-09-01
/ PRIOR APPLICATION NUMBER: 60/229,907
/ PRIOR FILING DATE: 2000-09-01
/ PRIOR APPLICATION NUMBER: 60/229,858
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 60
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 28
/ LENGTH: 431
/ TYPE: PRT
/ ORGANISM: Pantoea Stewartii
US-09-941-947A-28

```

Query Match 100.0%; Score 2231; DB 10; Length 431;

Query Match	Best Local Similarity	Pred. No.	4.4e-218;
100.0%	100.0%		

```

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 MCHERAVIA DDBBECUBAI ONT 20E1V89CHRVTFEEOHCKA1V7GSDIGFOTVGL0THP 60

1 MSHFAVIAPPPFSHVRALQNLAEVLVARGHRVTFQQHDCNALVIGSDIGFQIVGLQIHF 00

1 MSHEAVIAPPPFESHVRALONLAOELVARGHRVTFEQOHDCKALVTGSDIGFQTVGLQTHP 60

b 61 PGSLSHLLHAAHPLGSPMLRLINENWARTSDMLCRLPAALFALQIEGVIVDQMEPAGAV 120
y 121 VAEASGLPFVSVACALPLNREPGPLVAVMPFPGTSDAARERTTSEKIYDWMERHDRV 180
b 121 VAEASGLPFVSVACALPLNREPGPLVAVMPFPGTSDAARERTTSEKIYDWMERHDRV 180
y 181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPFRKALPCDFHAGVPLRQPGQTGSS 240
b 181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPFRKALPCDFHAGVPLRQPGQTGSS 240
y 241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
b 241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
y 301 GDIOVDFADQSAALSOAQLTTHGGMTVLDALASRTPELLALPLAFDOPGVASRIYVHG 360
b 301 GDIOVDFADQSAALSOAQLTTHGGMTVLDALASRTPELLALPLAFDOPGVASRIYVHG 360
y 361 IGRASRTTSHALARQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMRTQOP 420
b 361 IGRASRTTSHALARQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMRTQOP 420
y 421 VLSGQDYATAT 431
b 421 VLSGQDYATAT 431
RESULT 2
S-10-218-118-4
Sequence 4, Application US/10218118
Publication No. US20030148319A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: C11876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
PRIOR FILING DATE: 2002-08-13
PRIOR FILING DATE: 2002-08-13
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 431
TYPE: PRT
ORGANISM: Pantoea stewartii
S-10-218-118-4
Query Match 100.0%; Score 2231; DB 14; Length 431;
Best Local Similarity 100.0%; Pred No. 4,4e-218;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
y 1 MSHFAVIAPPPFSHVRLALQALVARGHRTVFFQOHDCKALVTGSDIGFTVGLQTHP 60
b 1 MSHFAVIAPPPFSHVRLALQALVARGHRTVFFQOHDCKALVTGSDIGFTVGLQTHP 60
y 61 PGSLSHLLHAAHPLGSPMLRLINENWARTSDMLCRLPAALFALQIEGVIVDQMEPAGAV 120
b 61 PGSLSHLLHAAHPLGSPMLRLINENWARTSDMLCRLPAALFALQIEGVIVDQMEPAGAV 120
y 121 VAEASGLPFVSVACALPLNREPGPLVAVMPFPGTSDAARERTTSEKIYDWMERHDRV 180
b 121 VAEASGLPFVSVACALPLNREPGPLVAVMPFPGTSDAARERTTSEKIYDWMERHDRV 180
y 181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPFRKALPCDFHAGVPLRQPGQTGSS 240
b 181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPFRKALPCDFHAGVPLRQPGQTGSS 240
y 241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300

Db 241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Qy 301 GDIOVDFADQSAALSOAQLTTHGGMTVLDALASRTPELLALPLAFDOPGVASRIYVHG 360
Db 301 GDIOVDFADQSAALSOAQLTTHGGMTVLDALASRTPELLALPLAFDOPGVASRIYVHG 360
Qy 361 IGRASRTTSHALARQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMRTQOP 420
Db 361 IGRASRTTSHALARQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMRTQOP 420
Qy 421 VLSGQDYATAT 431
Db 421 VLSGQDYATAT 431
RESULT 3
US-10-156-761-14886
Sequence 14886, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14886
LENGTH: 397
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14886
Query Match 11.3%; Score 251.5; DB 14; Length 397;
Best Local Similarity 26.6%; Pred. No. 1.2e-16;
Matches 115; Conservative 61; Mismatches 186; Indels 71; Gaps 18;
Qy 1 MSHFAVIAPPPFSHVRLALQALVARGHRTVFFQOHDCKALVTGSDIGFTVGLQTHP 60
Db 1 MGRFLFVVPPLVGHVNPVAVCTAAALARGHDIAWAGHPVLVGLAGADAVWFPCL---P 57
Qy 61 PGSLSHLLHAAHPLGSPMLRLINE--MARTSDMLCRLPAALFALQIEGVIVDQMEPAG 118
Db 58 EDGLSRPAGLK---GPAAFQFLWESFLVPLADAMAPGVRAAIEAYDPDVVYCDQAVAG 113
Qy 119 AVVAEASGLPFVSVACALPLNREPGPLVAVMPFPGTSDAARERTTSEKIYDWMERH 178
Db 114 ALVAESGLRPPWTSAT---SALVDPLAGMP-----KVAANL-----D 149
Qy 179 RVIAHACRM---GLA--PREKLHCHFCSPLAQISQILPELDPFRKALPCDFHAGV--LR 231
Db 150 GLLGELRRRTITGAGAADR-----FSPHGVLAVTTRALLGPVE--LPDRVWLVGPSVAA 202
Qy 232 OPQGTGSSITSFPSPDKPRIIPASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLS- 290
Db 203 RPAGPDOPFWMELEASALPTVLVSLGTANNDAGAPLNAA-----AEL-----GGIAD 251
Qy 291 ATQAGELARGDITQ-----VVDFAQSAALSOAQLTTHGGMTVLDALASRTPELLA 342
Db 252 RVRAVLVDPGVVEHPVDPDTVLVRRVYVQLALLERLDAVVCAGHTVCEALWHGVPLV 311
Qy 343 LPALFOPGVASRIYVHGIGKASRTTSHA--LARGISLITNTDYPQMTK-IQAALR 399
Db 312 APTRDQDPITVAQVVDAGAGVRL-RFRGADAARIAGAVBAVLDPAGGHRKAAAVGESFR 370

400 LAGCTPAAADIVE 412
|||:|||||:
371 AAGGSAADRL 383

SULT 4
-10-156-761-8483
Sequence 8483, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR FILING DATE: 2001-05-29
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-06-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8483
LENGTH: 412
TYPE: PRT
ORGANISM: Streptomyces avermitilis
-10-156-761-8483

Query Match 11.1%; Score 247; DB 14; Length 412;
Best Local Similarity 21.3%; Pred. No. 3.7e-16;
Matches 102; Conservative 76; Mismatches 162; Indels 138; Gaps 15;

3 HPVIAPIFFSHRALQNLAEIVARGHRTFFQCHDKKALVTGSDI-----49
4 HPLFMSAPFGHGWPPSLVAEELVHRGHVTF-----VTGAEDAVRSVGADFLR 54
50 ---GFTVGLQTHPPGSLHLLAAHPLGSMRLINEMA--RTSMLCRELPAAAFHAL 104
55 YESAFEGVDMYRLMTAEAPNAIPMTLYDEGMSMLRSVEEFGVDKVPDLVAYDIATSLN-- 112
105 QLEGVIVDQMEPAGVVAEASGLFPVSACALPLNRPGLPLAVMPFEGTSDAAREYT 164
113 -----VGRVLAASWSRPAMTY-----IPLF-----ASNGRFS 139
165 TSEKIYDMLRRHDRVIAHACRMGLAPREKLHHCFS---PLAQISQLIPEL-----213
140 TMQSVLD-----PDSAQVSAPPFR-----FSQOMELFGLGALVPRLEALLVSRG 183
214 -----DPPRKALPDCTH-----AVGPLRQPGTSGSSTSYFSPSPDK 249
184 ITEPVDDFLSGPEDFNLVCLPRAFOYAGDTDFERPAFVGPCLGKRGGLGWTP--PGSGH 241
250 PRIFASLGTQCHRYGLERTTAKACEEVDQAQLLAHCGGLSATQAGELARGD-----302
242 PVVLISLGTVFNQKSPFRIFVRAETDVPVHVVIS-----LGRGVDVLRPL 289
303 ---IQVDFADQSAALSQAQLTITHGGMNTVLDATASRTPLLLALPLAPDQPGVASRIVYH 359
290 PENVEVHRVPHHVALEHARALVTHGGTGSVMEALHAGCPVLVMPFSRDAQVTGRIAEAL 349
360 GICKRAS-RFTTSHALAQIRSLTNTDYPQGMTKIOAALRLAGCTPAAADIVEQAMR 416
350 GLGRMVQPEEVTATLLRRHVLDIISDDAITRCVQWQRATVEAGCALRAADETERFLR 407

35ULT 5
3-09-740-029-4
Sequence 4, Application US/09740029

Patent No. US20020082194A1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Kari et al
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
DRUG-METABOLIZING PROTEINS,
AND USES THEREOF
FILE REFERENCE: CL000928
CURRENT APPLICATION NUMBER: US/09/740,029
CURRENT FILING DATE: 2000-12-20
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 477
TYPE: PRT
ORGANISM: Oryctolagus cuniculus
US-09-740-029-4

Query Match 7.9%; Score 177; DB 9; Length 477;
Best Local Similarity 21.5%; Pred. No. 6.1e-09;
Matches 96; Conservative 75; Mismatches 177; Indels 98; Gaps 19;

QY 13 SHVRALQNLAEIVARGHRTFFQCHDKKALVTGSDIGFTQVGLQTHFP-----GSLSHL 67
DB 1 SHWINLVLEELQURGHRTVLPSP-SLLLDHTKIPFNVEVLQVTKETLMEEELNTV 59
QY 68 LHLAAHPGDSM-----LRLINEMWARTSDMLCRELPAAAFHALQIEG 108
DB 60 LYMSSFEL-PTLSWTKVLGKVMGKQPSKMLRRVCDSAITNKELLDLXAA-----KFDI 114
QY 109 VIVDQMEPAGVVAEASGLFPV-----SVACALPLNRPGLPLAVMPFET--GT 155
DB 115 CLADPLAFCGELVAELNIPVYSFRPSIGNIIRSCA-----GLP--TPSSYVPGS 164
QY 156 SDABERVTTSSEKIYDML-RRHDEVIAH-----ACR-MGLAPREK 195
DB 165 TSGUTDNMSFVQRLGNWLLYLMNDMFSHMLSWDEYYSKVLGRRTTICEIMCKAEWML 224
QY 196 LHCFSPLAQISQLIPELDPFRKALPDCFHAVGPL--RQPGTSGSSTSYFSPSP--DKPR 251
DB 225 IRSYV-----DEFFRPPLPN-FEYVGLHCKPAKPLPEELEEFVQSSGNDGW 272
QY 252 IFASLGTQCHRYGLFRTIAKACEEVDQAQLLAHCGGLSATQAGELARGDIOVVDFADQ 311
DB 273 VFTLGLSMIQNLTEERSNLINSALAQIPQKVLWRVTGKKPAT-----LGNTRLFEWIFQ 326
QY 312 SAALS--QAQLTITHGGMNTVLDATASRTPLLLALPLAPDQPGVASRIVYHIGIKRAS-RF 368
DB 327 NDLGHGPKTRAFITHGGTNGLYEALYHGVPMVGIPLFGDQPDNIARVKAKGAADVDLRI 386
QY 369 TTSALARQIRSLTNTDYPQGMTKI 394
DB 387 MTTSSLLKALKADVINNPSYKENAMKL 412

RESULT 6
US-09-962-678-2
Sequence 2, Application US/09962678
Patent No. US20020155499A1
GENERAL INFORMATION:
APPLICANT: Leiby, Kevin R.
TITLE OF INVENTION: 32624, A NOVEL HUMAN UDP-GLUCURONOSYL
TRANSFERASE AND USES THEREOF
FILE REFERENCE: 10448-094001
CURRENT APPLICATION NUMBER: US/09/962,678
CURRENT FILING DATE: 2001-09-25
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/235,044
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT

ORGANISM: Homo sapiens

US-09-962-678-2

Query Match 7.5%; Score 167; DB 9; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

6 VIAPPPFHVRLQNLQALQELVARGHRTVFPQDCKALV---TGSDIGFQTVGLQTHPPG 62

27 LWPCDMSHMLNWKVILEELIVGHEVTVL-TSKPSLIDYRKPSALKFEVVMQPDRT 85

63 SLSHLLHAAHPL-----GFSMLRLIN---EMARTSDMLC-----RELPAAFALQIEG 108

86 ENEIFVDLALNVLPLGLSTQSVIKLNDFFVEIRGTLKMMCESFIYQTLMKKLQETNYDV 145

109 VIVDQMPAGAVARASGLPFV---SVACALPLNREPG---LPLAVMPPEY-GTSDAARE 161

146 MLIDPVTIPCGDLMAELLAVFVLTLRISVCGNMERSCGKLPAPLSYVPVMTGLTD---- 201

162 RYTTSEKI-----YD-----WLMRHRDR 179

202 RMTFLERVKNMSLVLFHFHFIQDYDHFWEFYSKALGRPTTLCTETVGKAEIWLIRTY-- 259

180 VIAHACRMGLAPREKHLHCFSPLAQISQLIPDLDFPRKALPDCFHAVGEL--RQPGQTP 237

260 -----WDFEFPQYQPN-FEFVGLHCKHPAKALP 287

238 GSSTSYPPSPDKPRIPA-SLGTLL-QCHRYGLFRTIAKACEEVDQAQLLHACGGLSATQAG 295

288 KEMENFVQSSGEGDIIVFVSLGSLFQNVTEEKANIIASALAQIPQKVLNRYKGGKPS- 344

296 ELARGGDIQVDFADQSAALS--QAOLTIHGGNTVLDIAISRTPELLALPLAFDQ---- 349

345 ---LGNATRLYDWTIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMGVPIFGDQDNI 401

350 -----PGVASRIVVHGIGKGRASRTTSHALAROIRSLTNTDYPQMTKIQAAALRLA 401

402 AHMKAKGAAVEINF-----KTMISEDLLRALRTVITDSSYKEN-----AMRLS 444

RESULT 7

US-09-981-353-166

Sequence 166, Application US/09981353

Patent No. US20020160382A1

GENERAL INFORMATION:

APPLICANT: Lasek, Amy W.

APPLICANT: Jones, David A.

TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

FILE REFERENCE: PA-0038 US

CURRENT APPLICATION NUMBER: US/09/981.353

CURRENT FILING DATE: 2001-10-11

NUMBER OF SEQ ID NOS: 194

SOFTWARE: PERL Program

SEQ ID NO 166

LENGTH: 527

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. US20020160382A1 2434655CD1

US-09-981-353-166

Query Match

Best Local Similarity 7.5%; Score 167; DB 9; Length 527;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

6 VIAPPPFHVRLQNLQALQELVARGHRTVFPQDCKALV---TGSDIGFQTVGLQTHPPG 62

27 LWPCDMSHMLNWKVILEELIVGHEVTVL-TSKPSLIDYRKPSALKFEVVMQPDRT 85

63 SLSHLLHAAHPL-----GFSMLRLIN---EMARTSDMLC-----RELPAAFALQIEG 108

86 ENEIFVDLALNVLPLGLSTQSVIKLNDFFVEIRGTLKMMCESFIYQTLMKKLQETNYDV 145

PRIOR	FILING DATE: 1998-05-07	
PRIOR	APPLICATION NUMBER: 60/085573	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085579	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085580	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085582	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085700	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/086023	
PRIOR	FILING DATE: 1998-05-18	
PRIOR	APPLICATION NUMBER: 60/086392	
PRIOR	FILING DATE: 1998-05-22	
PRIOR	APPLICATION NUMBER: 60/086486	
PRIOR	FILING DATE: 1998-05-22	
PRIOR	APPLICATION NUMBER: 60/087098	
PRIOR	FILING DATE: 1998-05-28	
PRIOR	APPLICATION NUMBER: 60/087208	
PRIOR	FILING DATE: 1998-05-28	
PRIOR	APPLICATION NUMBER: 60/087609	
PRIOR	FILING DATE: 1998-06-02	
PRIOR	APPLICATION NUMBER: 60/087759	
PRIOR	FILING DATE: 1998-06-02	
PRIOR	APPLICATION NUMBER: 60/087827	
PRIOR	FILING DATE: 1998-06-03	
PRIOR	APPLICATION NUMBER: 60/088025	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088028	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088029	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088033	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088167	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088202	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088212	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088217	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088326	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088555	
PRIOR	FILING DATE: 1998-06-09	
PRIOR	APPLICATION NUMBER: 60/088722	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088738	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088740	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088811	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088824	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088825	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088826	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088861	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/088863	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/088876	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/089090	
PRIOR	FILING DATE: 1998-06-12	
PRIOR	APPLICATION NUMBER: 60/089105	
PRIOR	FILING DATE: 1998-06-12	
PRIOR	APPLICATION NUMBER: 60/089512	
PRIOR	FILING DATE: 1998-06-16	

PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089908

Query Match 7.5%; Score 167; DB 13; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;
Y 6 VIAPPPFHVRLQNLAEVLVARGHVTFFQOHDCKALV---TGSDIGFQTVGLQTHPPG 62
D 27 LVNFCDSHNLNKKVLEELIVGHEVTVL---THSKPSLDYRKPSALKFEVVMPODRTE 85
Y 63 SLSHLLHAAHPL-----GPSMLRLN---EMARTSDMLC-----RELPAAFHALQIEG 108
D 86 ENEIFVDLALNVLPGSLTWSQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145
Y 109 VVDQMPAGAVVAEASGLPFV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
D 146 MLIDPVPICGDLMAELLAVFVLTLRISVGGNMRSCGKLPAPLSVVPVMTGLTD---- 201
Y 162 RYTTSEKI-----YD-----WLMRHRDR 179
D 202 RMTFLERKVMKMSLVLFHFWIQDYVHFWEFYSKALGRPTTLCTVGVKAEIWLRTY-- 259
Y 180 VIAHACRMGLAPREKLHHCFSPLAQISQLIPBLDPRKALPDCHEAVGLP--RPOQGT 237
D 260 -----WDFEPFPQYQPN--PEFVGLHCKPAKALP 287
Y 238 GSSTSYFSPDKPRIFA-SLGTIL-QGHRYGLFRTIAKACEVDQAQLLHACGSLSATQAG 295
D 288 KEMENFVQSSGEGDIVVFSIGSLFQNVTEKANIISALAQIPQKVLWRYKKGKST--- 344
Y 296 ELARGGDIQVDFADQSAALS--QAQUTITHGWNVTVDALASRTPELLALPLAFDQ--- 349
D 345 ---LGANTRYLDYDIPQNDLLGHPKTKAFITHGGMNGIYEALYHGVPMVGVPIFGDQDNI 401
Y 350 -----PGVASRIVHVGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
D 402 AHMKAKGAARVEINF-----KMTTSEDLLRALRTVITDSSYKEN-----AMRLS 444

RESULT 9
US-10-174-590-522
Sequence 522, Application US/10174590
Publication No. US20030008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42
CURRENT APPLICATION NUMBER: US/10/174.590
CURRENT FILING DATE: 2002-06-18
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT

ORGANISM: Homo Sapien
US-10-174-590-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;
QY 6 VIAPPPFHVRLQNLAEVLVARGHVTFFQOHDCKALV---TGSDIGFQTVGLQTHPPG 62
DB 27 LVNFCDSHNLNKKVLEELIVGHEVTVL---THSKPSLDYRKPSALKFEVVMPODRTE 85
QY 63 SLSHLLHAAHPL-----GPSMLRLN---EMARTSDMLC-----RELPAAFHALQIEG 108
DB 86 ENEIFVDLALNVLPGSLTWSQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145
QY 109 VVDQMPAGAVVAEASGLPFV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
DB 146 MLIDPVPICGDLMAELLAVFVLTLRISVGGNMRSCGKLPAPLSVVPVMTGLTD---- 201
QY 162 RYTTSEKI-----YD-----WLMRHRDR 179
DB 202 RMTFLERKVMKMSLVLFHFWIQDYVHFWEFYSKALGRPTTLCTVGVKAEIWLRTY-- 259
QY 180 VIAHACRMGLAPREKLHHCFSPLAQISQLIPBLDPRKALPDCHEAVGLP--RPOQGT 237
DB 260 -----WDFEPFPQYQPN--PEFVGLHCKPAKALP 287
QY 238 GSSTSYFSPDKPRIFA-SLGTIL-QGHRYGLFRTIAKACEVDQAQLLHACGSLSATQAG 295
DB 288 KEMENFVQSSGEGDIVVFSIGSLFQNVTEKANIISALAQIPQKVLWRYKKGKST--- 344
QY 296 ELARGGDIQVDFADQSAALS--QAQUTITHGWNVTVDALASRTPELLALPLAFDQ--- 349
DB 345 ---LGANTRYLDYDIPQNDLLGHPKTKAFITHGGMNGIYEALYHGVPMVGVPIFGDQDNI 401
QY 350 -----PGVASRIVHVGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
DB 402 AHMKAKGAARVEINF-----KMTTSEDLLRALRTVITDSSYKEN-----AMRLS 444

RESULT 10
US-10-176-758-522
Sequence 522, Application US/10176758
Publication No. US20030008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176.758
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-758-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRLAQLAQLVARGHRVTFQDCKALV---TGSIDIGFTVGLQTHPPG 62
 b 27 LVWPCDMSHLNWKVLEELIVRGHEVTVL--THSKPSLDYKPSALKFVYVHPQDRTE 85
 Y 63 SLSHLLHLAAHPU-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108
 b 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145
 Y 109 VIVDQMEPAGVAAEASGLPFV---SVACALPLNREP---LPLAMPEY-GETSDAARE 161
 b 146 MLIDPVIPOCDLMAELLAVPFVTLIRISVGGNMERSCGKLPAPLSYVVPVMTGLTD----- 201
 Y 162 RYTTSEKI-----YD-----WLMRHRDR 179
 b 202 RMTFLERVNSMLSVLFHFHWIQQDYDHFWEFFYSKALGRPTTLCTETVGKAEIWLRTY-- 259
 Y 180 VIAHACRMGLAPREKLHCHFCPLAISOILIPELPPKALPDCEHAGPL--RPOQGTTP 237
 b 260 -----WDFBFPQPYQPN--PEFVGGLHCKPKAKALP 287
 Y 238 GSSTSYFPPSPDKPRIPA--SIGTL--QGHRYGLFRTTAKACEEVDQAOLLAHCGGLSATQAG 295
 b 288 KENENFVQSGGEGIVFVLSGLFQNVTEKANIISALAQIPQKVLWRYKCKKPEST--- 344
 Y 296 ELARGDIDQVPFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349
 b 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
 Y 350 -----PGVASRIYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401
 b 402 AHMKAGRAVEINF-----KTMTEDDLALRTVITDSSYKEN-----AMRLS 444

ESULT 11

S-10-175-737-522

Sequence 522, Application US/10175737

Publication No. US20030013153A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC50

CURRENT APPLICATION NUMBER: US/10/175,737

CURRENT FILING DATE: 2002-06-19

Prior Application removed - See file Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 522

LENGTH: 527

TYPE: PRT

ORGANISM: Homo Sapien

S-10-175-737-522

Query Match 7.5%; Score 167; DB 14; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRLAQLAQLVARGHRVTFQDCKALV---TGSIDIGFTVGLQTHPPG 62

b 27 LVWPCDMSHLNWKVLEELIVRGHEVTVL--THSKPSLDYKPSALKFVYVHPQDRTE 85

Y 63 SLSHLLHLAAHPU-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108

b 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145

Y 109 VIVDQMEPAGVAAEASGLPFV---SVACALPLNREP---LPLAMPEY-GETSDAARE 161

b 146 MLIDPVIPOCDLMAELLAVPFVTLIRISVGGNMERSCGKLPAPLSYVVPVMTGLTD----- 201

Y 162 RYTTSEKI-----YD-----WLMRHRDR 179

b 202 RMTFLERVNSMLSVLFHFHWIQQDYDHFWEFFYSKALGRPTTLCTETVGKAEIWLRTY-- 259

Y 180 VIAHACRMGLAPREKLHCHFCPLAISOILIPELPPKALPDCEHAGPL--RPOQGTTP 237

b 260 -----WDFBFPQPYQPN--PEFVGGLHCKPKAKALP 287

Y 238 GSSTSYFPPSPDKPRIPA--SIGTL--QGHRYGLFRTTAKACEEVDQAOLLAHCGGLSATQAG 295

b 288 KENENFVQSGGEGIVFVLSGLFQNVTEKANIISALAQIPQKVLWRYKCKKPEST--- 344

Y 296 ELARGDIDQVPFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349

b 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401

Y 350 -----PGVASRIYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401

b 402 AHMKAGRAVEINF-----KTMTEDDLALRTVITDSSYKEN-----AMRLS 444

QY 109 VIVDQMEPAGVAAEASGLPFV---SVACALPLNREP---LPLAMPEY-GETSDAARE 161
 Db 146 MLIDPVIPOCDLMAELLAVPFVTLIRISVGGNMERSCGKLPAPLSYVVPVMTGLTD----- 201
 QY 162 RYTTSEKI-----YD-----WLMRHRDR 179
 Db 202 RMTFLERVNSMLSVLFHFHWIQQDYDHFWEFFYSKALGRPTTLCTETVGKAEIWLRTY-- 259
 QY 180 VIAHACRMGLAPREKLHCHFCPLAISOILIPELPPKALPDCEHAGPL--RPOQGTTP 237
 Db 260 -----WDFBFPQPYQPN--PEFVGGLHCKPKAKALP 287
 QY 238 GSSTSYFPPSPDKPRIPA--SIGTL--QGHRYGLFRTTAKACEEVDQAOLLAHCGGLSATQAG 295
 Db 288 KENENFVQSGGEGIVFVLSGLFQNVTEKANIISALAQIPQKVLWRYKCKKPEST--- 344
 QY 296 ELARGDIDQVPFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349
 Db 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
 QY 350 -----PGVASRIYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401
 Db 402 AHMKAGRAVEINF-----KTMTEDDLALRTVITDSSYKEN-----AMRLS 444

RESULT 12

US-10-173-706-522

Sequence 522, Application US/10173706

Publication No. US20030022293A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCES: P3430RIC7

CURRENT APPLICATION NUMBER: US/10/173,706

CURRENT FILING DATE: 2002-06-17

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 522

LENGTH: 527

TYPE: PRT

ORGANISM: Homo Sapien

US-10-173-706-522

Query Match 7.5%; Score 167; DB 14; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

QY 6 VIAPFFSHVRLAQLAQLVARGHRVTFQDCKALV---TGSIDIGFTVGLQTHPPG 62

Db 27 LVWPCDMSHLNWKVLEELIVRGHEVTVL--THSKPSLDYKPSALKFVYVHPQDRTE 85

QY 63 SLSHLLHLAAHPU-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108

Db 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145

QY 109 VIVDQMEPAGVAAEASGLPFV---SVACALPLNREP---LPLAMPEY-GETSDAARE 161

Db 146 MLIDPVIPOCDLMAELLAVPFVTLIRISVGGNMERSCGKLPAPLSYVVPVMTGLTD----- 201

QY 162 RYTTSEKI-----YD-----WLMRHRDR 179

```

Db 202 RMTFLERVKMSLVLFHFWIQQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
Y 180 VIAHACRMGLAPREKLHCFSPALQISQLIPDLDFPRKALPDCFHAVGEL--RQPGGTP 237
Db 260 -----WDFEPQPYQPN--FEFVGLGCKPAKALP 287
Y 238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
Db 288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEKANIISALAQIPQKVLWRYGKKPST--- 344
Y 296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
Db 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGPPIFGDQDNI 401
Y 350 -----PGVASRIYVHGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
Db 402 AHMKAGAAVEINF-----KTMTSDDLRLRVTITDSSYKEN-----AMRLS 444

RESULT 13
US-10-175-738-522
Sequence 522, Application US/10175738
Publication No. US2003002294A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C45
CURRENT APPLICATION NUMBER: US/10/175,738
CURRENT FILING DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-738-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRALQNLQAEVLVARGHRTVFFQHDCKALV---TGSDIGFQTVGLQTHPPG 62
Db 27 LVWPCDMSHNLNVKVLLEELIVRGHEVTVL--THSKPSLIDYRKPSALKFEVVMPPQDRT 85
Y 63 SLSHLLHIAAHL-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108
Db 86 ENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGLTKMNCESPIYNTLMKGLQETNYDV 145
Y 109 VIVDQMEPAGAVAAEASGLFPV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
Db 146 MLIDFVPCGDLMAELLAVFPVLTLSVSGNMERSCKLPAPLSVYVPVMTGLTD----- 201
Y 162 RYTTSSEKI-----YD-----NLMRRHDR 179
Db 202 RMTFLERVKMSLVLFHFWIQQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
Y 180 VIAHACRMGLAPREKLHCFSPALQISQLIPDLDFPRKALPDCFHAVGEL--RQPGGTP 237
Db 260 -----WDFEPQPYQPN--FEFVGLGCKPAKALP 287
Y 238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
Db 288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEKANIISALAQIPQKVLWRYGKKPST--- 344
Y 296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
Db 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGPPIFGDQDNI 401
Y 350 -----PGVASRIYVHGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
Db 402 AHMKAGAAVEINF-----KTMTSDDLRLRVTITDSSYKEN-----AMRLS 444

```

```

Db 288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEKANIISALAQIPQKVLWRYGKKPST--- 344
Y 296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
Db 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGPPIFGDQDNI 401
Y 350 -----PGVASRIYVHGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
Db 402 AHMKAGAAVEINF-----KTMTSDDLRLRVTITDSSYKEN-----AMRLS 444

RESULT 14
US-10-175-752-522
Sequence 522, Application US/10175752
Publication No. US2003002295A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT APPLICATION NUMBER: US/10/175,752
CURRENT FILING DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-752-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRALQNLQAEVLVARGHRTVFFQHDCKALV---TGSDIGFQTVGLQTHPPG 62
Db 27 LVWPCDMSHNLNVKVLLEELIVRGHEVTVL--THSKPSLIDYRKPSALKFEVVMPPQDRT 85
Y 63 SLSHLLHIAAHL-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108
Db 86 ENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGLTKMNCESPIYNTLMKGLQETNYDV 145
Y 109 VIVDQMEPAGAVAAEASGLFPV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
Db 146 MLIDFVPCGDLMAELLAVFPVLTLSVSGNMERSCKLPAPLSVYVPVMTGLTD----- 201
Y 162 RYTTSSEKI-----YD-----NLMRRHDR 179
Db 202 RMTFLERVKMSLVLFHFWIQQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
Y 180 VIAHACRMGLAPREKLHCFSPALQISQLIPDLDFPRKALPDCFHAVGEL--RQPGGTP 237
Db 260 -----WDFEPQPYQPN--FEFVGLGCKPAKALP 287
Y 238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
Db 288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEKANIISALAQIPQKVLWRYGKKPST--- 344
Y 296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
Db 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGPPIFGDQDNI 401

```

350 -----PGVASRIYVHGIGKRASRFTTSHALARQIRSLTNTDYPORMTKIQALRLA 401
402 AHMKAKGAARVEINF-----KWTMSDILLRALRTVITDSSYKEN-----AMRLS 444

RESULT 15

-10-176-482-522

Sequence 522, Application US/10176482

Publication No. US20030022296A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C70

CURRENT APPLICATION NUMBER: US/10/176,482

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 522

LENGTH: 527

TYPE: PRT

ORGANISM: Homo Sapien

-10-176-482-522

Query Match

Best Local Similarity 20.8%; Score 167; DB 14; Length 527;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

6 VIAPFFSHVRALQNLQELVARGHRVTFQOHDCKALV---TQSDIGFQTVGLQTHPPG 62

27 LVWPCDMSHMLNVKVLSELIVRGHEVTVL-TSKPSLIDYRKPSALKFEVWHMPQDRTE 85

63 SLSELHLAAAPL-----GPSMLRLIN---EMARTSDMLC-----RELPAAPHALQIEG 108

86 ENEIFVDLAINVLPGLSWQSVIKLNDFFVEIRGLTKMWCESFIYNQTLMKKLOETNYDV 145

109 VIVDQMEFAGAVAEASGLPFV---SVACALPLNREPG---LPLAVMPFEY-GTSDAARE 161

146 MLIDSVIFCGDLMAELLAVPFVLTIRISVGNMERSCKLPAPLSYFVPEVMTGLTD---- 201

162 RYTTSEKI-----YD-----MLRRHR 179

202 RMTFLERVNKMVLVLFHFWIQDDYHFWEFYSKALGRPTTLCTGVKAEIWLIRTY-- 259

180 VIAHACRMGLAPREKLHCFSPLAQISQILPELDFPRKALPDCFHAVGEL--RQPGGTP 237

260 -----WDZEPQPYQPN-FEFVGGGLCKPAKALP 287

238 GSSTSYFSPDKPRIFA-SLGTL-QCHRYGLFRIAKACEVDAQILLAHCGSLSATQAG 295

288 KEMENFVQSSGSDGIVWFSLFQNTTEKANTIASALAQIPOKVLRYRKKPST--- 344

296 ELARGDITQWVDFADQSAALS--QAQLTITHGQNTVLDAIASRTPLIALPLAFDQ---- 349

345 ---LGNTRLYDWIFQNDLLGHPTKAPITHGKNGNIVEALYHGVPMVGVPIFGDQDNI 401

350 -----PGVASRIYVHGIGKRASRFTTSHALARQIRSLTNTDYPORMTKIQALRLA 401

402 AHMKAKGAARVEINF-----KWTMSDILLRALRTVITDSSYKEN-----AMRLS 444

Search completed: February 29, 2004, 15:27:58

Job time : 32.6159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 11.9967 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947a-28
Project score: 2231
Sequence: 1 MSHFAVIAPPPFHVRLQN.....EQAMRTQPVLSGQDYATAL 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 78*

1: PIR1*

2: PIR2*

3: PIR3*

4: PIR4*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1794	80.4	431	2 S52583	ctx protein - Erw
2	1790	80.2	431	2 B37802	ctx protein - Erw
3	1018.5	45.7	413	2 S52980	zeaxanthin glucosyl
4	558.5	25.0	419	1 S74500	zeaxanthin glucosyl
5	262	11.7	407	2 T37104	probable glycosyl
6	228.5	10.2	430	1 S33184	glycosyltransferase
7	217	9.7	424	2 T51110	glycosyltransferase
8	204	9.1	392	1 C69851	macrolide glycosyl
9	198.5	8.9	418	2 F75587	probable glycosyl
10	195.5	8.8	287	2 H69306	macrolide glycosyl
11	181.5	8.1	418	2 J50636	glycosyl transferase
12	180	8.1	54	2 A46292	zeaxanthin glucosyl
13	175.5	7.9	541	2 JCS423	2-hydroxyacylshpin
14	175	7.8	407	2 C97234	probable glycosyl
15	172	7.7	388	2 B70878	probable transferase
16	168.5	7.6	611	2 T03890	hypothetical prote
17	167	7.5	455	1 XUBHFG	flavonol 3-O-gluc
18	164.5	7.4	541	2 A48801	2-hydroxyacylshpin
19	164	7.4	392	2 S72936	UDP-glucuronosyltr
20	162	7.3	535	2 I57961	glucuronosyltransf
21	160.5	7.2	449	2 G70670	hypothetical prote
22	160	7.2	533	2 T27589	probable glycosyl
23	158	7.1	414	2 D70723	probable transferase
24	154.5	6.9	391	1 A69784	macrolide glycosyl
25	149	6.7	428	2 C70670	probable glycosyl
26	148.5	6.7	471	1 S08325	flavonol 3-O-gluc
27	147.5	6.6	395	2 AB0242	probable glycosyl
28	147.5	6.6	530	2 JCS656	UDP glucuronosyltr
29	146.5	6.6	530	2 S17512	glucuronosyltransf

hypothetical prote
flavonol 3-O-gluc
flavonol 3-O-gluc
glucuronosyltransf
probable glycosyl
glucuronosyltransf
glucuronosyltransf
glycosyltransferase
hypothetical prote
glucuronosyltransf
probable glycosyl
hypothetical prote
hypothetical prote
hypothetical prote
glucuronosyltransf
hypothetical prote

ALIGNMENTS

RESULT 1

S52583

ctx protein - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999

C:Accession: S52583

R:Lin, Y.P.; Lai, B.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52583

A>Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 1-431 <LIN>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21261.1; PID:G148395

C:Superfamily: glycosyltransferase

Query Match 80.4%; Score 1794; DB 2; Length 431;

Best Local Similarity 82.4%; Pred. No. 3.1e-132;

Matches 355; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

Qy	1	MSHFAVIAPPPFHVRLQNLAQELVARGHRTVTFQOHDCKALVTGSDIGFQVGLQTHP	60
Db	1	MSHFAIAPFFSHVRLAENLAQELVARGHQTFTFQQOQTKPLIOHPHIGFHAVGESSHP	60
Qy	61	PGSLSHLLHAAHPGLGPMRLINEMWARTSDMLCKELPAAFFALQIEGVIVDQMEPAGAV	120
Db	61	SGSLSLHLLHAAHPGLGPMRLINEMWARTSDMLCLALPRAFNDAIDGVIVDQMEPAGAL	120
Qy	121	VAPASGLPFVSVACALPLNRPGLPLAVMPFPGTSDAARERTTSEKIYDWLVRHDRV	180
Db	121	VAEALGLPFVSVACALPLNRPGLPLAVMPFDYGTSDAARERTTSEKIYDWLVRHDRV	180
Qy	181	IAHACGEMGLAPREKLHCFSPALQISOLIPELDTPRKALPDCFFHVGPLRQPGTGPSS	240
Db	181	IAHSHRMGLAPRDLKHLQCFSPALQISOLIPELDTPRKALPACFFHVGPLRQTSTSS	240
Qy	241	TSYFSPDKPRIFASLGTGLQCHRYGLFTIAKACBEVDAQULLAHCGGLSATQAGELARG	300
Db	241	PLYFSPDKPRIFASLGTGLQCHRYGLFKAIVKACEIDQQLLAHCGRLTAFQAEELARS	300
Qy	301	GDIOVDFDQSNALSOAQTLTHGGMTVLDAISRTPLLALPLAFDQPGVASIVVHG	360
Db	301	RFTQVVDFDQSAALSOAQLAJTHGGMTVLDAINRYRTPLLALPLAFDQPGVASIVVHG	360
Qy	361	IGKRASRFTTSHALARIQIRSLTNTDYPQMTKIQAALFACGTPAAADIVSQAMRTQCP	420
Db	361	IGKASRFTTSHALAHMALLFNVDYPMQMTKIQTALFLAGGTAAADIVSQAMRTQCP	420
Qy	421	VLGGQDYATAL 431	
Db	421	VLKGGYATAL 431	

```

Query Match      45.7%; Score 1018.5; DB 2; Length 413;
Best Local Similarity 48.8%; Pred. No. 8.8e-72;
Matches 211; Conservative
Qy 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGF---QTVGLQ 57
Db 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGF---QTVGLQ 57
Qy 58 THP-PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEP 116
Db 58 THP-PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEP 116
Qy 52 FYPLPASVQQAQRNVQOQSMENLLRLIAAMSSLDVLCQOLPAILOQLAVDALIVDEMPP 111
Db 52 FYPLPASVQQAQRNVQOQSMENLLRLIAAMSSLDVLCQOLPAILOQLAVDALIVDEMPP 111
Qy 117 AGAVVAAAGLPPVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLR 176
Db 117 AGAVVAAAGLPPVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLR 176
Qy 112 AGSLVAEALGPPFISIAACALPVNREPGCLPLVNPFPFYAEDKRALRFQVSEIYDALMTP 171
Db 112 AGSLVAEALGPPFISIAACALPVNREPGCLPLVNPFPFYAEDKRALRFQVSEIYDALMTP 171
Qy 177 HDRIYAHACRMGLAPREKLHCFSPLAQISQLIPLOFPFKALPDCFHAGVPLRPOQT 236
Db 177 HDRIYAHACRMGLAPREKLHCFSPLAQISQLIPLOFPFKALPDCFHAGVPLRPOQT 236
Qy 172 HGQTLRHAQRFGLPERRLOECLSPLAQISQSPALDPFRALPNCFYVGAIR----- 226
Db 172 HGQTLRHAQRFGLPERRLOECLSPLAQISQSPALDPFRALPNCFYVGAIR----- 226
Qy 237 PGSSTSYFESP-----DKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGL 289
Db 237 PGSSTSYFESP-----DKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGL 289
Qy 227 -----YQPPQVSPSPRSTPRIFASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGL 280
Db 227 -----YQPPQVSPSPRSTPRIFASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGL 280
Qy 290 SATQAGELARGGDIQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQ 349
Db 290 SATQAGELARGGDIQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQ 349
Qy 281 TPAQADSLVACGATEVVSFVDPQRYVAEANLVTHGGLNTVLDALAAATFVLAVPLSPDQ 340
Db 281 TPAQADSLVACGATEVVSFVDPQRYVAEANLVTHGGLNTVLDALAAATFVLAVPLSPDQ 340
Qy 350 PGVASRIYVHGIGKASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAAD 409
Db 350 PGVASRIYVHGIGKASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAAD 409
Qy 341 PAVARLVYVNGLRVSRFARQOTLADETLAQLGLDTHQRIATARQQLNDAGGTPPRAAT 400
Db 341 PAVARLVYVNGLRVSRFARQOTLADETLAQLGLDTHQRIATARQQLNDAGGTPPRAAT 400
Qy 410 IVEQAMTQCPV 421
Db 410 IVEQAMTQCPV 421
Qy 401 LIEQAIAGESV 412
Db 401 LIEQAIAGESV 412

RESULT 4
S74500
zeaxanthin glucosyl transferase crtX - Synecocystis sp. (strain PCC 6803)
N:Alternate names: hypothetical protein slr1125
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74500
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-419 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651250; PIDN:BAAL6652.1; PID:g16517;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: crtX
C:Superfamily: glycosyltransferase

Query Match      25.0%; Score 558.5; DB 1; Length 419;
Best Local Similarity 31.7%; Pred. No. 6.8e-36;
Matches 132; Conservative 81; Mismatches 196; Indels 7; Gaps 5;
Qy 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGFQTVGLQ 60
Db 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGFQTVGLQ 60
Qy 61 PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEPAGA 119
Db 61 PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEPAGA 119
Qy 61 LGAQAEFMAELGKLSIKALQITVAKITOKRAAFAPEAPGVAKAGVAVLLVDQVSQEGG 120
Db 61 LGAQAEFMAELGKLSIKALQITVAKITOKRAAFAPEAPGVAKAGVAVLLVDQVSQEGG 120

Query Match      80.2%; Score 1790; DB 2; Length 431;
Best Local Similarity 80.5%; Pred. No. 6.3e-132;
Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;
Qy 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGFQTVGLQ 60
Db 1 MSHFAVIAIPFPFVSHVRALQNLAEIVARGHRVTFPQOHDCKALVTGSDIGFQTVGLQ 60
Qy 61 PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEPAGAV 120
Db 61 PGSLSHLLHAAHPLGSPMLRLINEMARTSDMLCRELPAAFPALQIEGVTVDMQEPAGAV 120
Qy 121 VAPASGLPPVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLRHDRV 180
Db 121 VAPASGLPPVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLRHDRV 180
Qy 121 VAEALGLPPIVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLRHDRV 180
Db 121 VAEALGLPPIVSVACALPLNREPGCLPLAVMPPEYGTSDAARERTTSEKIYDLWLRHDRV 180
Qy 181 IAHSHRMGLAPREKLHCFSPLAQISQLIPLOFPFKALPDCFHAGVPLRPOQTGSS 240
Db 181 IAHSHRMGLAPREKLHCFSPLAQISQLIPLOFPFKALPDCFHAGVPLRPOQTGSS 240
Qy 241 TSYPSPDKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Db 241 TSYPSPDKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Qy 241 SRYTSPDKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Db 241 SRYTSPDKPRIPASIGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Qy 301 GDIQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQPGVASRIYVHG 360
Db 301 GDIQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQPGVASRIYVHG 360
Qy 301 RHTQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQPGVASRIYVHG 360
Db 301 RHTQVVDPADQSAALSOAQLTTHGGMNTVLDIASRTPPLALPLAFDQPGVASRIYVHG 360
Qy 361 IKGASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAADIVEQAMTQCP 420
Db 361 IKGASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAADIVEQAMTQCP 420
Qy 361 IKGASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAADIVEQAMTQCP 420
Db 361 IKGASRFTTSHALARQIRSLTNTDYPQRTKIOAALRLAGGTPPAADIVEQAMTQCP 420
Qy 421 VLSGGDYATAT 431
Db 421 VLSGGDYATAT 431
Qy 421 VLSGGDYATAT 431
Db 421 VLSGGDYATAT 431

RESULT 3
S52980
zeaxanthin glucosyl transferase - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999
R:Hundle, B.; Alberti, M.; Nievelstein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Bur
Mol. Gen. Genet. 245, 406-416, 1994
A:Title: Functional assignment of Erwinia herbicola Eho 10 carotenoid genes expressed in
A:Reference number: S52976; MUID:95107236; PMID:7808389
A:Accession: S52980
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <HUN>
A:Cross-references: EMBL:M87280; NID:g148404; PIDN:AAAG4979.1; PID:g148410
C:Superfamily: glycosyltransferase
```


A: Molecule type: DNA
A: Residues: 1-424 <QUI>
A: Cross-references: EMBL:AF055579; PIDN: AAC12648.1
A: Experimental source: ATCC 11891
A: Genes: yjC
A: Superfamily: glycosyltransferase
A: Description: intracellular inactivation of oleandomycin by glycosylation [validated, M
A: Note: very specific for oleandomycin
A: Keywords: glycosyltransferase, hexosyltransferase
Query Match 9.7%; Score 217; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 3e-09;
Matches 111; Conservative 69; Mismatches 185; Indels 84; Gaps 19;
QY 3 HFAVIAPFFSHVRLQNLQAEVLVARGHRTFFQOHDCKALV-----TGSDFGFTVGLQ 57
DB 14 HISFPNIPGCHVNPSTLALVEKLCEKGRVTVATTEFAFAPVQOAGGALIVYTSLNID- 67
QY 58 THPPGSLHLHLAAHPLGSMRLINEMARTSDMLC---RELPAAPFALQIEG---VTV 111
DB 68 -----ILPEEFNEBELLAEDQGRWACSLAEAPVLPQLRTATPTTGRLDIV 115
QY 112 DQMPAGAVVAESGLPFV-----SVACA-----LPLNREPGPL---AVMPFEGVTSDA 158
DB 116 DIASWAPVLGRKWDIPFVQLSPSTVAEYGFESDPAVQDPTADRGEEAAAP---ACTGDA 173
QY 159 ARRYTTSEKIYDLMERHDRVTA---HH-----ACHMGLAPREKLHCFSPLAQCSQLI 210
DB 174 --EGAGAE---DGLVFFTRLSAFLEHGVDPFATFLIAPNRCIVGCRAP---SQIK 224
QY 211 PELDFPRKALPDCFHAVGPL---RQPGT---PSSSTVSFPPDKPRIIPASLGTLQGHRY 264
DB 225 GD-----TVGDNVTFVGPVYDGRSHQGTWEGPHG-----RPLVLLIALGSAFTDHL 270
QY 265 GLFTIAKACEVDQAQLLHACGLSLATQAGELARGDIOVDFADQSAALSOAQLTIH 324
DB 271 DFYATCLSAVDGLDHWVLSVGRFVDPADLGEVP---PNVEVHQWVQPOLDILTKASAFITH 328
QY 325 GGMNVLDAISRPPLALPLAFDQPGVASRIYVHGIGKRAAS-FTTSHALARAQIRSLFT 383
DB 329 AGMSTWELNSVNPVAVVQIAEQTNWAEIVELGLGRHIPDQVTAEKLEAVLAVAS 388
QY 384 NTDYPORMKTIQAAALRAGGTPAAADIVE 412
DB 389 DPGVAERLAAVQRIEAGGARAADILE 417
RESULT 8
A: Molecule type: DNA
A: Residues: 1-424 <QUI>
A: Cross-references: EMBL:AF055579; PIDN: AAC12648.1
A: Experimental source: ATCC 11891
A: Genes: yjC
A: Superfamily: glycosyltransferase
A: Description: intracellular inactivation of oleandomycin by glycosylation [validated, M
A: Note: very specific for oleandomycin
A: Keywords: glycosyltransferase, hexosyltransferase
Query Match 9.7%; Score 217; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 3e-09;
Matches 111; Conservative 69; Mismatches 185; Indels 84; Gaps 19;
QY 3 HFAVIAPFFSHVRLQNLQAEVLVARGHRTFFQOHDCKALV-----TGSDFGFTVGLQ 57
DB 14 HISFPNIPGCHVNPSTLALVEKLCEKGRVTVATTEFAFAPVQOAGGALIVYTSLNID- 67
QY 58 THPPGSLHLHLAAHPLGSMRLINEMARTSDMLC---RELPAAPFALQIEG---VTV 111
DB 68 -----ILPEEFNEBELLAEDQGRWACSLAEAPVLPQLRTATPTTGRLDIV 115
QY 112 DQMPAGAVVAESGLPFV-----SVACA-----LPLNREPGPL---AVMPFEGVTSDA 158
DB 116 DIASWAPVLGRKWDIPFVQLSPSTVAEYGFESDPAVQDPTADRGEEAAAP---ACTGDA 173
QY 159 ARRYTTSEKIYDLMERHDRVTA---HH-----ACHMGLAPREKLHCFSPLAQCSQLI 210
DB 174 --EGAGAE---DGLVFFTRLSAFLEHGVDPFATFLIAPNRCIVGCRAP---SQIK 224
QY 211 PELDFPRKALPDCFHAVGPL---RQPGT---PSSSTVSFPPDKPRIIPASLGTLQGHRY 264
DB 225 GD-----TVGDNVTFVGPVYDGRSHQGTWEGPHG-----RPLVLLIALGSAFTDHL 270
QY 265 GLFTIAKACEVDQAQLLHACGLSLATQAGELARGDIOVDFADQSAALSOAQLTIH 324
DB 271 DFYATCLSAVDGLDHWVLSVGRFVDPADLGEVP---PNVEVHQWVQPOLDILTKASAFITH 328
QY 325 GGMNVLDAISRPPLALPLAFDQPGVASRIYVHGIGKRAAS-FTTSHALARAQIRSLFT 383
DB 329 AGMSTWELNSVNPVAVVQIAEQTNWAEIVELGLGRHIPDQVTAEKLEAVLAVAS 388
QY 384 NTDYPORMKTIQAAALRAGGTPAAADIVE 412
DB 389 DPGVAERLAAVQRIEAGGARAADILE 417
RESULT 9
A: Molecule type: DNA
A: Residues: 1-418 <WHI>
A: Cross-references: EMBL:AF001863; GB:AE001825; NID:g6460670; PIDN:AAF12451.1; PID:g646074
A: Experimental source: strain R1
A: Genes: yjC
A: Superfamily: glycosyltransferase
A: Map position: 2
A: Description: 8.9%; Score 198.5; DB 2; Length 418;
A: Best Local Similarity 22.9%; Pred. No. 8.1e-08;
A: Matches 98; Conservative 79; Mismatches 210; Indels 41; Gaps 15;
QY 6 VIAFPFFSHVRLQNLQAEVLVARGHRTFFQOHDCKALVTSGLI---GFTVGLQTEPPG 62
A: Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A: Reference number: A69580; MUID:98044033; PMID:9384377
A: Accession: C6981
A: Status: preliminary
A: Molecule type: DNA

5 IASQPIAGHVLPUREIARELARRGHELRWYTGKRFQVQEAAGAVWEGF--VHARDYDDA 62

63 SLSHLLHLAAHGLPMSMLRLINEMARTSOMLCRELPAAPHAL-----QIEGVIVOME 115

63 NFD-----AAPGERSARRGLAQLLFDLRHVFVQMGEGQFYDLRLERARWRDPVWLADQTV 117

116 PAGAVVAASGLPFFVSACALPL---NREP---GLPLAVMPFEVGTSDAARER--YTTSE 167

118 GA-ALLREELGPPVALLGLVPLGFIASRDYAPFGGLAPLG---GAQOLNRALQVLITE 173

168 KIYDMLRRHRVIAHACRMGLAPREKLHCHFCSPLOAISOLIPELDPPKALPDCPHAV 227

174 RV---VFRGISQLADLCTWGL-PAFSFAPVAPVAPSLMQLPVSERLEYLRDLPPQVRFI 229

228 GPLROPOG--TPGSSTVSPSPDKPRIIPASLGTGLQGHRYGLFRTIAKACEVDQAOLLAH 285

230 GLPLVPAGEFTPPAWNLDILTRPVVLTQGLATREQLRPAALMALADEDLVVAA- 288

286 CGLSATQAGELARGDQIOWDFADOSALSOAQLTTHGGMNTVLDIASRTPLALPL 345

289 --GVAPERLGLPL--ANARAAAFVFTALLPRHVASVLTNGGVTQVQALSHGVPCVWAGR 344

346 AFDPQGVASRIYVHGIGKXASFTTSHALAR-QIRSLTNTDYQRMTKIOAALRLAGGT 404

345 SEDKAEVAAVAYSAGLNLGTATSPARIAAVRVUREPAFREQAGHGLGHELRQHDAP 404

405 PAAADIVE 412

405 READWLE 412

35006

acridole glycosyltransferase homolog yojK - Bacillus subtilis

Species: Bacillus subtilis

Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

Accession: H69906

Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Geller, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Ketter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadaie, Y.; Sato, T.; Scanlon, Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron, Teuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Banchin, A.

Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

Reference number: A69580; PMID:98044033; PMID:9384377

Accession: H69906

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

Residues: 1-287 <KUN>

Cross-references: GB:Z99114; GB:AL009126; NID:G2634230; PIDN:CAB13834.1; PID:G2634335

Experimental source: strain 168

Genetics:

Gene: yojK

Superfamily: glycosyltransferase

Query Match 8.8%; Score 135.5; DB 2; Length 287;

Best Local Similarity 26.5%; Pred. No. 8.5e-08;

Matches 44; Conservative 46; Mismatches 73; Indels 3; Gaps 2;

252 IPASLGTQGHRYGLFRTIAKACEVDQAOLLAHCGGLSATQAGELARGDQIOWDFADQ 311

116 LFIISGTLFNNQKQFNQCKVDFDKGKVLISGKIKTSELNDIPE--NFIVRPVVPQ 173

312 SAALSOAQLTTHGGMNTVLDIASRTPLALPLAFDPQGVASRIYVHGIGKXASFTTSH 371

Db 174 LEILKRASLFTVTHGGMNSTSEGLYFETPLVLPWPGDQFVWADQVEKVGAGKVKKEELS 233

QY 372 HALARQ-IRSLTNTDYQRMTKIOAALRLAGGTTPAAADIVEQAMR 416

Db 234 ESLKKEITQEVMMNRSYAKAKXBIGQSLKAGGSKKAADSTLEAVK 279

RESULT 11

JS0636

glycosyl transferase (EC 2.4.1.1-) - Streptomyces lividans

C:Species: Streptomyces lividans

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999

C:Accession: JS0636

R:Jenkins, G.; Cundliffe, E.

Gene 103, 55-62, 1991

A:Title: Cloning and characterization of two genes from Streptomyces lividans that confere

A:Reference number: JS0635; PMID:92104506; PMID:1761231

A:Molecule type: DNA

A:Residues: 1-418 <JEN>

A:Cross-references: GB:M74717; NID:G153345; PIDN:AAA26780.1; PID:G153347

A:Experimental source: strain TK21

C:Comment: This enzyme transfers glucose from UDP-glucose to macrolide antibiotics.

C:Genetics:

A:Gene: mgt

C:Superfamily: glycosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 8.1%; Score 181.5; DB 2; Length 418;

Best Local Similarity 22.7%; Pred. No. 1.7e-06;

Matches 101; Conservative 67; Mismatches 186; Indels 91; Gaps 17;

QY 2 SHFAVIAPPFFSHVRALONLAQELVARGHRTVTFQOHDCKKALVTGSDIGTQVGLQTHPP 61

Db 23 AHTAMPSIALHGHVNPVLEIVRELVARHRTVTAIPRLADKV--AEAGAEPKLMNSTLP 80

QY 62 GSLSHLHLAHP--LGPMSRLINEMARTSOMLCRELPAAPHALQIEGVIVDQMEPAGA 119

Db 81 GP-----DADPRWGSTLLDNVZPFLDALQSLPQLAQYEGDEPLVLEDIASYAR 133

QY 120 VVAASGLPFFVSA-CALP---LNPPEGLPLAVMPFEVGTSDAARERTTSEKIYDMLAR 175

Db 134 VLGRREWEVPIVSLSPVAVWEGVEGVEGPEPMWEPKTERGOVYVARE-----HAML-- 185

QY 176 RDRVIAHACRMGLAPREKLHCHFCSPLOAISOLIPELDPPKALPDCPHAVGPLROPOG 235

Db 186 EENGITDHPDPFGIPGPRSLV-----LIP-----KAL-----QPHA 216

QY 236 TPGSSTVSP-----PSPDKPRIIPASLGTQGHRYGLFRTIAKACEVD 278

Db 217 DRVDETYYTFVACCGQRTAGDWARPGEAKVVLVSLGSAFTKQPAFYRECVRAGFLP 276

QY 279 AQLLAHCGGLSATQAG---ELARGCD---IQWDFADQSAALSOAQLTTHGGMNTVL 331

Db 277 -----GMHTVLQVGRHVDPAELGDVDPNVVETWVPQALILQQADLFVTHAGAGSQ 328

QY 332 DALASRTPLALPLAFDPQGVASRIYVHGIGKXASFTTSHALARQIR-----SLITNTDY 387

Db 329 EGLATATPMIAVPQADQFGNADML--QGLGV-ARTLPTTEATATAKALTAALALVDDDEV 385

QY 388 PORTMKIOAALRLAGGTTPAAADIVE 412

Db 386 AARLKEIQARMAQEAQTRGPADLIE 410

RESULT 12

A46292

zeaxanthin glucosyltransferase - Erwinia herbicola (fragment)

C:Species: Erwinia herbicola

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Aug-1998

C:Accession: A46292

R:Hundle, B.S.; O'Brien, D.A.; Alberti, M.; Beyer, P.; Hearst, J.E.

Proc. Natl. Acad. Sci. U.S.A. 89, 9321-9325, 1992

Title: Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola*
Reference number: A46292; MUID:93028456; PMID:1409639
Accession: A46292
Status: preliminary; not compared with conceptual translation
Molecule type: nucleic acid
Residues: 1-54 <HUN>
Note: sequence extracted from NCBI backbone (NCBIP:115898)
Superfamily: glucosyltransferase

Query Match 8.1%; Score 180; DB 2; Length 54;
Best Local Similarity 61.1%; Pred. No. 1.6e-07;
Matches 33; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

306 VDFADQSAALSQAQLTHGGMNTVLDIASRTPLALPLAFDQPCVASRIYVH 359
1 VSFVDQPVYVAEANLVTHGGINTVLDALAAATVLAPELSPDQPAVARLVYN 54

RESULT 13
C5423
-hydroxyacylphosphingosine 1-beta-galactosyltransferase (EC 2.4.1.45) - human
Alternate names: UDP-galactose:ceramide galactosyltransferase
Species: Homo sapiens (man)
Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 26-Aug-1999
Accession: J05423
Kapitonov, D.; Yu, R.K.
Biochem. Biophys. Res. Commun. 232, 449-453, 1997
Title: Cloning, characterization, and expression of human ceramide galactosyltransferase
Reference number: J05423; MUID:97242209; PMID:9125199
Accession: J05423
Molecule type: mRNA
Residues: 1-541 <KAP>
Cross-references: GB:U62899
Experimental source: fetal glioma cell
Comment: This enzyme catalyzes the final step of galactosylceramide synthesis.
Genetics:
Gene: cgt
Superfamily: glucuronosyltransferase
Keywords: glycoprotein; glucosyltransferase; hexosyltransferase
472-492/Domain: hydrophobic #status predicted <HYD>
539-540/Region: endoplasmic reticulum retention signal #status atypical
78,333,442/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 175.5; DB 2; Length 541;
Best Local Similarity 20.1%; Pred. No. 7e-06;
Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;

6 VIAPP--FFSHVRALONLAQELVARGHRTFFQHDCKALVTGSDIG-FQTGLQTHP-- 60
24 IIVPPIMFESHMTVEKTLASALHERGHTVFL-----LSEGRDIAPSNHYSLQRYPGI 76
61 -----PGSLSHLLHLAAHPLGPSMLRLINEMARTSDMLCRELPAAFH 102
77 FNSTTSDAFLQSKMRNIFSGRLTAI-----ELFDLIDHYTKNDMM-----VGNH 121
103 ALQTEG-----VIYDQMEPAGAVAAEAGLPFVSACALPLNREPG--LPVAMPPE 152
122 AL-IQGLKKEKFDLLVDPNDCGFFVLAHLGVKXAVFSTGLVYPAEYVAGAPALAYVP-E 179
153 YGTSDAABERYTTSEKIYDLMRRHRDVIHAHACWCLA-----PREKLHCHFSPLAQISQ 208
180 FNSLLTQMN-----LLOGMKVTGYLSRGLVSVPLVPK-----YERIMQKN 223
209 LIPE-----LDFPRXALPDCPAGVGLPQPGTQPGSSTSPSPSPD 248
224 LPEKSMYDLVHGSSLMWLTCDVALEPRPTLPNIVYVGGILTKPAS-----PLPE 274
249 KPRIPASGLTQGHRYGLFRTIAKACEVDAQLLAHCGSLATQAGELAR----- 299
275 DLQRVANGANEHGFVLPVPGAGVYLSIEDIANKL-----AGALGRUPQKVIWRF 323
300 -----GGSIQVVDVADQSAALSQAQLT--ITHGGMNTVLDIASRTPLALPLAFDQ 349

324 SGPKPKNLGNNTKLIWLPQNDLLGHSKIKAFVSHGGLNSIFETWYHGVVGVGVFGDH 383
350 PGVASRIVVHIGIKRAS-RFTTSHALARQIRSLTNTDYPQRTKIQAALRLAGGTP 405
384 YDTWTRVQAKGNGILLENKVTTEKELYBALVKVINNPESYRQRAQKLSEIHKQPGHP 440

RESULT 14
C97234
probable glycosyl transferase from UDP-glucuronosyltransferase family CAC2716 [imported]
Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C/Accession: C97234
R.Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: C97234
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-407 <KUR>
A/Cross-references: GB:AB001437; PIDN:AAK80662.1; PID:g15025750; GSPDB:GN00168
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC2716

Query Match 7.8%; Score 175; DB 2; Length 407;
Best Local Similarity 20.0%; Pred. No. 5.3e-06;
Matches 89; Conservative 69; Mismatches 178; Indels 108; Gaps 14;

10 PFFSHVRALONLAQELVARGHRTFFQHDCKALVTGSDIGFQTVGLQTHPQSLSHLLH 69
10 PGYCHVNPTGLVDLKVGEVYVCTDFKESIEAAGAKPISYGER-----MSNFKN 63
70 LAAHPLGPSMLRLINEMARTSDMLCRELPAAFHALQIEGVIVDQMEPAGAVAAEASGLPF 129
64 ANKTAGNRSYADITTKIIGTCEIEDILKIKGMKFDYVVMHCSMFCAGNIIAQILKVP 123
130 VSVACALPLNREPGLPVAMPPEYG-----TSDAARERTTSEKIYDMLMRH 177
124 ISFAPVATPKZ-----VNMQYSPITEKLESNIVSRAYSEAKRIEKKYFVRMPAL 177
178 DRVIAHACR--NGLAPREKLHCHFSPLAQISOLIDELQPPKALPDCPFAHVP--LRQP 233
178 SELM--ACRGDLNIAATK-----EFVSNIEY---YDPSFKFTGPPYIKRK 218
234 QQTSGSSTSYEP--SPDKPIEFASIGTL---QGHRYGLFRTIAKACEVDAQLLAHC 286
219 ENLD-----FPPEKLKNGKVIYISLGTVPNTNLSNLYNIF----- 253
287 GELSATOAGELARGGDIQVY-----DFADQSAALSQAQLTTHGG 326
254 -----PEAPGSDVVMVMSAYNIDTSNFINPKNFIVRNVYVQTEILKYADAATHAG 305
327 MNTVLDIASRTPLALPLAFDQPCVASRIYVHIGIKRASR-FTTSHALARQIRSLTNT 385
306 MNSTDLIYINKYFPVPIGADQLYMASRAKELGAAIVLNKQKLAKELRESYKVMVDA 365
386 DYPQRTKIQAALRLAGGTTPAAD 409
366 SYLENIEKIRDVFEAGGYKGA 389

RESULT 15
B70878
probable transferase - Mycobacterium tuberculosis (strain H37RV)
Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: B70878
R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

re 393, 537-544, 1998
thors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
tle: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome
ference number: A70500; MUID:98295987; PMID:9634230
cession: B70878
atus: preliminary; nucleic acid sequence not shown; translation not shown
ecule type: DNA
sidues: 1-388 <COL>
oss-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15535.1; PID:g262426
erimental source: strain H37Rv
netics:
ne: RV2739c
perfamily: glycosyltransferase

ery Match 7.7%; Score 172; DB 2; Length 388;
st Local Similarity 25.6%; Pred. No. 8.6e-06;
tches 90; Conservative 40; Mismatches 169; Indels 52; Gaps 12;

88 RTSDDLCLRELPAAFHALQIEGVIVDQMEPAGAVVAEASGLPFVSVACALPINREP-----142
72 RAAQMAVLNVP-RLRALEPELVVSDVITACGGAAELGIPWVE-----LNPHPLYLPS 124

143 -GLPLAVMPEYGTSDAAASERYTTSKIYDNLRRHDRVIAHACRMGLAPREKLHCFS 201
125 KGLPPICSGLAAGTGIRLRDNTMBALGTRSWRAGLRQRAAVRVEIGLPAEDP-----G 179

202 PLAQISQLPELDFPKALPDCFHAVGPLRQPGCTPGSSTSYFSPDKPRIPASLGTLOQ 261
180 FLRLIATLPALEVPDPWPAEAVVVGPL---HFEPTDRVLAIPAGTGPVVVWVAPSTALT 236

262 HRYGLFRTIAKACEEYDQALLAHCGGLSATQAGELA---RGGDIQVYDFA-----DQS 312
237 GTAGL-----TEVALQSLTP---GETVPSGSLVWSRLSGADLTVPFWAVAGLGSQA 285

313 AALSQAQLTITHCGMNTVLDIASRTPLALPLAFDQPGVAGRIYVHGIGKASRTTSH 372
286 ELLTRADLVICGGCHGWAKTLLAGVPMVVPVGGGQDWEIANRVVRQG-SAVLIRPLTAD 344

373 ALARQIRSLTNTDPCRMTKIQALRLAGGTFAADIVEQAMRTCPVLS 423
345 ALVAANVEVLSPRFR-----EARRAASVAGAADPV-----RVCHDALA 385

ch completed: February 29, 2004, 14:52:45
time : 14.9967 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:35:44, Search time 13.0973 Seconds
(without alignments)
1698.885 Million cell updates/sec

File: US-09-941-947A-28
Effect score: 2231
Sequence: 1 MSHRAVIAPPFSSVRALQN.....EQAMRTCPVLGGDYATAL 431

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1790	80.2	431	1	US-07-783-705A-2
2	960	43.0	399	1	US-08-096-623A-20
3	171.5	7.7	414	1	US-09-337-913-1
4	171.5	7.7	414	2	US-08-750-524-1
5	160.5	7.2	452	4	US-08-311-731A-6
6	142.5	6.4	426	3	US-09-320-878-8
7	142.5	6.4	426	3	US-09-105-537-20
8	142.5	6.4	426	4	US-09-141-908-8
9	142.5	6.4	426	4	US-09-657-440-8
10	142.5	6.4	3782	3	US-09-105-537-4
11	142	6.4	463	4	US-08-311-731A-7
12	138.5	6.2	530	4	US-09-356-806-113
13	138	6.2	531	5	PCT-US92-00282-6
14	137.5	6.2	408	2	US-08-926-258-2
15	137.5	6.2	408	2	US-09-120-053-2
16	137.5	6.2	408	3	US-09-198-212-2
17	137.5	6.2	408	4	US-09-789-261-2
18	137.5	6.2	431	2	US-08-576-628A-54
19	133	6.0	533	5	PCT-US92-00282-3
20	132.5	5.9	408	2	US-08-924-847A-2
21	132.5	5.9	408	3	US-09-120-052-2
22	132	5.9	409	2	US-08-924-254-2
23	132	5.9	409	3	US-09-120-249-2
24	129	5.8	529	5	PCT-US92-00282-7
25	126	5.6	524	4	US-09-356-806-40
26	123	5.5	528	4	US-09-356-806-8
27	119.5	5.4	455	3	US-09-036-987A-17

28	119.5	5.4	455	3	US-09-370-700-17
29	119.5	5.4	455	4	US-09-603-207-17
30	116.5	5.2	454	4	US-09-813-918-2
31	115.5	5.2	445	4	US-09-252-931A-17629
32	115	5.2	534	5	PCT-US92-00282-4
33	114.5	5.1	407	2	US-08-926-327-2
34	114.5	5.1	407	3	US-09-119-918-2
35	112	5.0	443	1	US-08-660-765A-4
36	111	5.0	390	3	US-09-036-987A-8
37	111	5.0	390	3	US-09-370-700-8
38	111	5.0	390	4	US-09-603-207-8
39	111	5.0	1209	4	US-09-252-991A-25844
40	108.5	4.9	530	3	US-09-180-852-2
41	108.5	4.9	740	4	US-09-624-693A-15
42	108	4.8	452	4	US-09-284-768A-22
43	108	4.8	608	4	US-09-284-768A-4
44	108	4.8	657	4	US-09-284-768A-7
45	106.5	4.8	531	5	PCT-US92-00282-5

ALIGNMENTS

RESULT 1
US-07-783-705A-2
Sequence 2, Application US/07783705A
Patent No. 5429939
GENERAL INFORMATION:
APPLICANT: Misawa, No. 5429939ihiko
APPLICANT: Kobayashi, Kazuo
APPLICANT: Nakamura, Katsumi
APPLICANT: Yamano, Shigeyuki
TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
SYNTHESIS OF CAROTENOID
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ladax & Parry
STREET: 26 West 61 Street
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10023
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 19-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 431 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-2

Query Match 80.2%; Score 1790; DB 1; Length 431;

Best Local Similarity 80.5%, Pred. No. 1.7e-178;
Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;

1 MSHFAVIAAPFFSHVRALQNLQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTH 60
1 MSHFAAIAAPFFSHVRALQNLQALVARGHRVTFIQYDIKHLIDSETIGFHSVGTDSHP 60

61 PGSLSHLHLAHPGLPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIVDMEPAGAV 120
61 PGALTRVHLAHPGLPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIVDMEPAGAL 120

121 VAERASGLPFFSVACALPLNREPGLPLAVMPEFFYGTSDAARERTTTSEKIYDMLMRHDRV 180
121 VAERAGLPFFSVACALPLNREPGLPLAVMPEFFYGTSDAARERTTTSEKIYDMLMRHDRV 180

181 IAHAACRMGLAPREKLHCHSPALQISOLIPDLDPFKALPDCFHAVGFLRQPGTSGS 240
181 IAESHREMGAPROKLGKCFSPALQISOLIPDLDPFKALPDCFHAVGFLRTHAPSTSS 240

241 TSYFPPKPRIFASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGLSATQAGELARG 300
241 SRYTTSSEKPRIFASLGTQCHRYGLFRTIVKACEEIDGQLLAHCGRLTDSQCEELARS 300

301 GDIOVDFADQSAALSQAQITTHGGMNTVLDALASRTPLLAIPALPFDGCVASRIYVHG 360
301 RHTQVDFADQSAALSQAQITTHGGMNTVLDALASRTPLLAIPALPFDGCVASRIYVHG 360

361 IGKASRFTTSHALAROIRSLTNTDYPQRTKIQAALRLAGGTAAADIVEQAMTCOP 420
361 IGKASRFTTSHALAROIRSLTNTDYPQRTKIQAALRLAGGTAAADIVEQAMTCOP 420

421 VLSGGDYATAL 431
421 VLSGGDYATAL 431

RESULT 2
US-08-096-623A-20
Sequence 20, Application US/08096623A
Patent No. 5684238
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L.
APPLICANT: Brinkhaus, Friedhelm L.
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Hwei-Che B.
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096.623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-20

Query Match 43.08; Score 960; DB 1; Length 399;
Best Local Similarity 48.6%; Pred. No. 9.5e-32;
Matches 204; Conservative 58; Mismatches 124; Indels 34; Gaps 7;

QY 1 MSHFAVIAAPFFSHVRALQNLQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQ 57
DB 1 MSHFAVIAAPFFSHVRALQNLQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQ 57

QY 58 THP-PGSLSHLLHLAHPGLPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIVDMEP 116
DB 52 FYPLPASVQOQRNVQOQNSGNLRLIAAMSSLDVLCCQLPALORLAVDALIVDEMEP 111

QY 117 AGAVVAEASGLPFFSVACALPLNREPGLPLAVMPEFFYGTSDAARERTTTSEKIYDMLMR 176
DB 112 AGSIVAEALGLPFFSIACALPVNRE--LPLVWPFHYAEDKRAAEQVSEIRYDALWYP 169

QY 177 HDRVIAHACRMGLAPREKLHCHSPALQISOLIPDLDPFKALPDCFHAVGFLRQPGT 236
DB 170 HGQTLRAQAFGLPERRLDCLSPALQISQSVPALDPFRALPNCFTYVVALR----- 224

QY 237 PGSSTSVFSP-----DKPRIPASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGL 289
DB 225 -----YQPPFQVSPSPRTPRIFASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGL 278

QY 230 SATQAGELARGDIIQVYDFADQSAALSQAQITTHGGMNTVLDALASRTPLLAIPALPFDQ 349
DB 279 TPAQADSLYCGA-TEVVSFVDQPRYVAEANLVITHGGLNTVLDALAAATPVLAVPLSFDQ 337

QY 350 PGVASRIYVHGIGKASRFTTSHALAROIRSLTNTDYPQRTKIQAALRLAGGTAAAD 409
DB 338 PAVAAALVYNGLGRVSRFARQOTLADETIAQLLQGLDETLHORVATASQQLNDAGTFCGD 397

RESULT 3
US-09-337-913-1
Sequence 1, Application US/09337913
Patent No. RE37206
GENERAL INFORMATION:
APPLICANT: Katsuhisa KOJURI, et al.
TITLE OF INVENTION: GENE ENCODING GLYCOSYLTRANSFERASE AND ITS
TITLE OF INVENTION: USES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.2
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/337,913
FILING DATE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/337,913
FILING DATE:
FILING DATE: December 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2401
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
I-09-337-913-1

Query Match 7.7%; Score 171.5; DB 1; Length 414;
Best Local Similarity 21.4%; Pred. No. 2.9e-09;
Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
6 VIAPFFSHVRALQNLAEVLVARGHRTF-----34
7 VATTFGDGHVNMVFAQEMVSRGHEVWYTKAPRSTVETGARHEPMRDAHDFGGMPR 66
35 ---FQHDCKALVTGSDIGFOTVGLQTHPPGSLHLLHAAHPGLSPMLRLINEMARTSD 91
67 EAPFQHAGLTCITGMIAGRDIFIE-----PAADQMTDLL-----102
92 MLCRELPAFAHALQIEGVIVDQMEPAGAVABASGLPFVSVACAL-----PLNREP 142
103 ALLEDFFA-----DVLVTDFTFGAGFVSERTGIPVAMIATSIYVFSRDTPAPLGL-- 153
143 GLPLAVMPEYCTSDAARERTTSKIDML-----MRSHDRVIAHACRMGLAPREKLHH 198
154 GLPPS-----SSRLGRNLTVLKQLTRVMDLRRAHVVD-----RVGL-PRIR-KG 201
199 CFSPLAQISQL-----IPELDPRKALPDCFHAGPLRQPGTSGSTSYFPPSD--KPR 251
202 AFENIMRTPDLVLLGTVPSEFYPRGDMPEVRFVGPVSPAPPDFTPPAMWGELDSGRPV 261
252 IPASLGTLQCHRYGLFRFAKACEEVDQALLAHCGGLSNTQAGELARGG-----301
262 VEVTTQ-----TVANDAE-----RLLLPAIRALAEDVLVWATTGAPLELEPMPA 306
302 DIQWVDFADQSAALQAOLITTHGGMNTVLDIAASRTPLLALPLAFDPQGVASRIVYHGI 361
307 NVYVERFIPHALLPHVDAMWYNGYGVNTALAHGVPLV-VARTEEKHEVAARVWSGA 365
362 GKR-ASRFTTSHALAQIRSLTNTDYPQ 389
366 GVHLKKRRLSERDIRRAVAVL---DEPR 391

RESULT 4
S-08-750-524-1
Sequence 1, Application US/08/50524
Patent No. 5861293
GENERAL INFORMATION:
APPLICANT: Katsuhisa KOJIRI, et al.
TITLE OF INVENTION: GENE ENCODING GLYCOSYLTRANSFERASE AND ITS
TITLE OF INVENTION: USES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSER: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
City: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.2
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,524
FILING DATE: December 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-750-524-1

Query Match 7.7%; Score 171.5; DB 2; Length 414;
Best Local Similarity 21.4%; Pred. No. 2.9e-09;
Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
6 VIAPFFSHVRALQNLAEVLVARGHRTF-----34
7 VATTFGDGHVNMVFAQEMVSRGHEVWYTKAPRSTVETGARHEPMRDAHDFGGMPR 66
35 ---FQHDCKALVTGSDIGFOTVGLQTHPPGSLHLLHAAHPGLSPMLRLINEMARTSD 91
67 EAPFQHAGLTCITGMIAGRDIFIE-----PAADQMTDLL-----102
92 MLCRELPAFAHALQIEGVIVDQMEPAGAVABASGLPFVSVACAL-----PLNREP 142
103 ALLEDFFA-----DVLVTDFTFGAGFVSERTGIPVAMIATSIYVFSRDTPAPLGL-- 153
143 GLPLAVMPEYCTSDAARERTTSKIDML-----MRSHDRVIAHACRMGLAPREKLHH 198
154 GLPPS-----SSRLGRNLTVLKQLTRVMDLRRAHVVD-----RVGL-PRIR-KG 201
199 CFSPLAQISQL-----IPELDPRKALPDCFHAGPLRQPGTSGSTSYFPPSD--KPR 251
202 AFENIMRTPDLVLLGTVPSEFYPRGDMPEVRFVGPVSPAPPDFTPPAMWGELDSGRPV 261
252 IPASLGTLQCHRYGLFRFAKACEEVDQALLAHCGGLSNTQAGELARGG-----301
262 VEVTTQ-----TVANDAE-----RLLLPAIRALAEDVLVWATTGAPLELEPMPA 306
302 DIQWVDFADQSAALQAOLITTHGGMNTVLDIAASRTPLLALPLAFDPQGVASRIVYHGI 361
307 NVYVERFIPHALLPHVDAMWYNGYGVNTALAHGVPLV-VARTEEKHEVAARVWSGA 365
362 GKR-ASRFTTSHALAQIRSLTNTDYPQ 389
366 GVHLKKRRLSERDIRRAVAVL---DEPR 391

RESULT 5
US-08-311-731A-6
Sequence 6, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESS: WOLF, GREENFIELD & SACKS, P. C.
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELEPHONE: 617/720-2441
TELEFAX: 617/720-3500
INFORMATION FOR SEQ ID NO: 6:
LENGTH: 452 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM TUBERCULOSIS

US-08-311-731A-6

Query Match 7.2%; Score 160.5; DB 4; Length 452;

Best Local Similarity 19.3%; Pred. No. 4.7e-08;

Matches 93; Conservative 67; Mismatches 154; Indels 169; Gaps 18;

9 PPFVSHVR-----ALONLAQELVARGHRTFFQOHCCKALVTGSDIGFQTVGLQTHPG 62

14 PPVASEVGRDAAISTTDAAPAGLAARRRILFVAE-----AVT----- 51

63 SLSLHLH--LAAPHPLGSMRLINEMARTSDMLCRLPAAAFHAL----- 104

52 LAHVVPFALAQSLDPSRYEVHVFACDPRYNQLLGLPFRHAIHTIPSRFFGNLTQGR 110

105 -----QIEGVIVDQMEP-----AGAVVAEASGLPFVSVACALPLNREPGL 144

111 FYAMRTLRKVEADLRVLDSEIADPLVVGDLRISLSVSARLAGIPYIAANAY----- 162

145 PLAVMPPEYGTSDAAREYTTSEKIY-----DWMKMR 176

163 -----WSPYAQRFFPLPDVITWTFGLVRLVLLYSLRPLLFALQCMPLNVRRR 212

177 HD-RVIAHACRMGLAPREKLHCFSPLAQISQILPELDPKALPDCHAVCP-LRQPO 234

213 HGLSSGLNLCRIPTDGDHLY-----ADVPMLMPTD-----LPANHEYGLPVLSWPA 261

235 GTPGSSSTYFSPDPKPRIFASLGTLOGHRYGLERTIAKACEEVDQALLAHCGGLSATOA 294

262 GKPTWDSLPT-DRPIVYATLGTSGGNL-----LQVLNALAELPVTVI 306

295 GELARGDIO-----VVFADQSAALSAQTLTIHGQNTVLDIASRTPLALPLAF 347

307 AATAARGDLKTVPANAFVADYLPGEAAARSAVVVCGGSLTTQQALVAGVPVI----- 360

348 DQPGVASSRIVH-----GIGKASRFTTSHALAQIRSLTNTDYPQRMKIQ 396

361 ---GVAGNLDQHLNWEAVERAGAGVLLRTSL-KSQRVAGAVMQVSRSEYFTSRATR 416

QY 397 ALR 399

Db 417 CLR 419

RESULT 6

US-09-320-878-8

; Sequence 8, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARSONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; EARLIER FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 8

; LENGTH: 426

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-320-878-8

Query Match 6.4%; Score 142.5; DB 3; Length 426;

Best Local Similarity 23.1%; Pred. No. 3.3e-06;

Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

QY 13 SHVRALONLAQELVARGHRTFFQOHCCKALVTGSDIGFQTVGLQTHPGSLSHLLH--- 69

Db 12 THYGLVPLAWALLAGHEVYVASQALTTTIGS-----GLAAVPGT-DHLIHEYR 63

QY 70 --LAAPHPLGSMRLINEMARTSDMLCRLPAAAF-HALQIEGVI-----VDQ 113

Db 64 VRMAGEP-RPNHPALAPDEARPS-----PLDWDHALGTETAILAPFYLLANNDSMVDD 115

QY 114 M-----EP---AGAVVAEASGLPFVSVACALPLNREPGLPLAVMPPEYGT 156

Db 116 LVDPARSQWDLVWETTYAGAAVQVGAARVLWG-----PDVWG--- 159

QY 157 DAARERTTSEKIYDMLMRSHDRVIAHACRMGLAPREKLHCFSPLAQISQILPELD--- 214

Db 160 -SARXKX-----VALDRQPPKHEDEPTAE--WLTWTLDRY 192

QY 215 ---PPRKALPDCHAVCP---LRQPGTSGSSTSYFP-----SPDKPRIFA 254

Db 193 GASPEEELLTQGF-TIDPTPPSLRLDTGLFTVGNRVYFNGTSVVPDMLSEPPARPVCL 251

QY 255 SIGYTL-----QCHRYGLPRTIAKACEEVDQALLAHCGGLSATQAGELAR--GGDIQV 306

Db 252 TLGVASAEVLGSDGVSGQ---DILEALADLDIELVAT---LDASQRAEIRNYPKHTRFT 304

QY 307 DFADQSAALSAQTLTIHGQNTVLDIASRTPLALPLAFDQPGVASRIVYHIGIKRAS 366

Db 305 DFVPMHALLPSCSAIHHGGAGFTYATAVINAVPQVLMELMDAPVKARVAEQAG--- 360

QY 367 RFTTSHALAQ-IRSLLTN-TDVPQRMKIQALLRAGGTAAADIVEQAMR 416

b 361 FFLPPAELTQAVRDVAVRLLDDPSVATAAHRRLRETTGDPPTAGIVPELER 412

RESULT 7

S-09-105-537-20

Sequence 20, Application US/09105537A

Patent No. 6265202

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 606.438US1

CURRENT APPLICATION NUMBER: US/09/105,537A

CURRENT FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 20

LENGTH: 426

TYPE: PRT

ORGANISM: Streptomyces venezuelae

S-09-105-537-20

Query Match 6.4%; Score 142.5; DB 3; Length 426;

Best Local Similarity 23.0%; Pred. No. 3.3e-06;

Matches 107; Conservative 42; Mismatches 191; Indels 125; Gaps 22;

Y 13 SHVRALQNLAEVLVARGHRVTFPQQHDKALVTGSDIGFTVGLQTHPPGSLSLHLH--- 69

b 12 THYGLVPLAWALLAAGHEVRVASQPALTDITGS-----GLAAPPVGT-DHLIHYR 63

Y 70 --LAAHPLGSMRLINEMARTSDM-----LCRELPAAFHALQIEGVVDQM----- 114

b 64 VRNAGEP-RPNHPAIAFDEARPEPLDWDHALGIEAILAPYFHLANNDMSWDDLVDFARS 122

Y 115 -----EP-----AGAVVABASGLPPVSVACALPLNRBGLPLAVMPFEYGTSDAAREY 163

b 123 WQDLVLEPTTYAGAAQVTAAGAAHVLMG-----PDVVG-----SARRKF 165

Y 164 TTSEKIYDMLRRHRDVIHAHACRMGLAPREKLHCFSPLAQISOLIPELD-----PPRK 218

b 166 -----VALDRQPPHREDPTAB--WLTWLDRYGASPEE 199

Y 219 ALPDCFHAVGP-----LRPQGTGPGSTSYFP-----SPDKPRIFASLGLT-- 259

b 200 LITGQP-TIDPTPPSLRLDTGLPTVMRYVPYNGTSVVPDWLSEPPARPRVCLTLGVSAR 258

Y 260 -----QCHRYGLPRTIAKACEVEDAQLLAHCGGLSATQAGELAR--GCDIQWDFADQSA 313

b 259 EVLGGDVSQG--DILEALADLDIELVAT----LDASQRAEIRNYPKHTRFTDFVPMHA 311

Y 314 ALSQAQLTTHGGMNTVLDIASRTPLLALPLAFDPQGVASRIVVHVGIGKASRFTTSHA 373

b 312 LLPSCSAIIHHGAGTYATAVINAVPQVMAELWDAPVKARAVAQGAG----PFLPPAB 367

Y 374 LARQ-IRSLTN-TDYPQMTKIQALRLAGCTPAADIVQAMR 416

b 368 LTPQAVRDVAVRLLDDPSVATAAHRRLRETTGDPPTAGIVPELER 412

RESULT 8

S-09-141-908-8

Sequence 8, Application US/09141908

Patent No. 6503741

GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary

APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, Mary

APPLICANT: MCDANIEL, Robert

APPLICANT: TANG, Li

TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a

; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold

; FILE REFERENCE: 300622002100

; CURRENT APPLICATION NUMBER: US/09/141,908

; CURRENT FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: PROV. 60/076,919

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: PROV. 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 426

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-141-908-8

Query Match 6.4%; Score 142.5; DB 4; Length 426;

Best Local Similarity 23.1%; Pred. No. 3.3e-06;

Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

QY 13 SHVRALQNLAEVLVARGHRVTFPQQHDKALVTGSDIGFTVGLQTHPPGSLSLHLH--- 69

Db 12 THYGLVPLAWALLAAGHEVRVASQPALTDITGS-----GLAAPPVGT-DHLIHYR 63

QY 70 --LAAHPLGSMRLINEMARTSDMLRELPAAF-HALOIEGVI-----VDQ 113

Db 64 VRNAGEP-RPNHPAIAFDEARPE-----PLDWDHALGIEAILAPYFHLANNDMSWDD 115

QY 114 M-----EP-----AGAVVABASGLPPVSVACALPLNRBGLPLAVMPFEYGT 156

Db 116 LVDFAFSWQDLVLEPTTYAGAAQVTAAGAAHVLMG-----PDVVG-- 159

QY 157 DAAREYTTSEKIYDMLRRHRDVIHAHACRMGLAPREKLHCFSPLAQISOLIPELD-- 214

Db 160 -SARRKF-----VALDRQPPHREDPTAB--WLTWLDRY 192

QY 215 ---PPRKALPDCFHAVGP-----LRPQGTGPGSTSYFP-----SPDKPRIFA 254

Db 193 GASFEELLTGQP-TIDPTPPSLRLDTGLPTVMRYVPYNGTSVVPDWLSEPPARPRVCL 251

QY 255 SLGTL-----QCHRYGLPRTIAKACEVEDAQLLAHCGGLSATQAGELAR--GCDIQV 306

Db 252 TLGVSAREVLGGDVSQG--DILEALADLDIELVAT----LDASQRAEIRNYPKHTRFT 304

QY 307 DFADQSAALSQAQLTTHGGMNTVLDIASRTPLLALPLAFDPQGVASRIVVHVGIGKAS 366

Db 305 DFVPMHALLPSCSAIIHHGAGTYATAVINAVPQVMAELWDAPVKARAVAQGAG---- 360

QY 367 RFTTSHALAQ-IRSLTN-TDYPQMTKIQALRLAGCTPAADIVQAMR 416

Db 361 FFLPPAELTQAVRDVAVRLLDDPSVATAAHRRLRETTGDPPTAGIVPELER 412

RESULT 9

US-09-657-440-8

; Sequence 8, Application US/09657440

; Patent No. 6509455

GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/657,440

; CURRENT FILING DATE: 2000-09-07

; PRIOR APPLICATION NUMBER: 09/320,878

; PRIOR FILING DATE: 1999-05-27

PRIOR APPLICATION NUMBER: CIP OF 09/141,908
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 8
LENGTH: 426
TYPE: PRT
ORGANISM: Streptomyces venezuelae

IS-09-657-440-8

Query Match 6.4%; Score 142.5; DB 4; Length 426;
Best Local Similarity 23.1%; Pred. No. 3.3e-06;
Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

2Y 13 SHVRALQNLAEVARGHRVTFQCHDCKALVTGSDIGFQTVGLQTHPPGSLHLH--- 69
DB 12 THYGLVPLAWALLAGHEVRVASQPALTDITGS-----GLAAVPVGT-DHLIHEYR 63
2Y 70 --LAAHPLGSMRLINEMARTSDMLCRELPAP-HALQIEGV-----VQO 113
DB 64 VRMAGEP-RPNHPAIAFDEARPEPLWDHALGIEATLAPYFYLLANNDMSVDD 115
2Y 114 M-----EP-----AGAVVABASGLPFVSVACALPLNREPGLPLAVMPFEGTS 156
DB 116 LVDFARSHQPLVLEWPTTAVAGAAVQTGAARVVLWG-----PDVWG-- 159
2Y 157 DAARERTTSEKIYDMLRRHDRVIAHACRMGLAPREKHLHCFSPALQISQILPELD-- 214
DB 160 -SARRKF-----DILEALADLDIELVAT-----LDASORAEIRNYPKTRFT 192
2Y 215 ---PFRKALPCFHAVGP---LRQPGTSGSSTSYEP-----SPDKPRIFA 254
DB 193 GASFEELTQGF-TIDTPPSLELDGLPTVGMRYVYNGTSVVPDMLSEPPARPVCL 251
2Y 255 SLGTL-----QGHRYGLFRTIAKACEEVDQAQLLAHCGLSATQAGELAR--GGDIQV 306
DB 252 TLGVSAREVLGGDGVSGQ---DILEALADLDIELVAT-----LDASORAEIRNYPKTRFT 304
2Y 307 DFADOSALSOAQLTIHGKNTVLDALASRTPLALPLAFDOPGVASRIVYHIGKRASRFTTSHA 366
DB 305 DFVPMHALLPSCSAIIHGGAGTATAVINAVPOVMLAELWDAPVKARVAEQAG----- 360
2Y 367 RFTTSHALARQ-IRSLTN-TDYPQRTMKIOAALRLAGGTPAAADIVEQAMR 416
DB 361 PFLPPAELTPQAVRDAVRIILDDPSVATAAHRLEETFGDPTPAGIVPELER 412

RESULT 10
US-09-105-537-4
Sequence 4, Application US/09105537A
Patent No. 6265202

GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3782
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-4

Query Match 6.4%; Score 142.5; DB 3; Length 3782;
Best Local Similarity 23.0%; Pred. No. 0.0001;
Matches 107; Conservative 42; Mismatches 191; Indels 125; Gaps 22;

2Y 13 SHVRALQNLAEVARGHRVTFQCHDCKALVTGSDIGFQTVGLQTHPPGSLHLH--- 69

DB 414 THYGLVPLAWALLAGHEVRVASQPALTDITGS-----GLAAVPVGT-DHLIHEYR 465
QY 70 --LAAHPLGSMRLINEMARTSDMLCRELPAP-HALQIEGV-----VQO 114
DB 466 VRMAGEP-RPNHPAIAFDEARPEPLWDHALGIEATLAPYFYLLANNDMSVDDIVDFARS 524
QY 115 -----EP-----AGAVVABASGLPFVSVACALPLNREPGLPLAVMPFEGTSDAARERY 163
DB 525 WQPDVLWLEWPTTAVAGAAVQTGAARVVLWG-----PDVWG--SARRKF 567
QY 164 TTSEKIYDMLRRHDRVIAHACRMGLAPREKHLHCFSPALQISQILPELD-----PFRK 218
DB 568 -----VALDRQPPHEHREDPTAE---WLTWTLDRYGASFEE 601
QY 219 ALPDCFHAVGP---LRQPGTSGSSTSYEP-----SPDKPRIFA 259
DB 602 LLTQGF-TIDTPPSLELDGLPTVGMRYVYNGTSVVPDMLSEPPARPVCLTLGVSAR 660
QY 260 -----QGHRYGLFRTIAKACEEVDQAQLLAHCGLSATQAGELAR--GGDIQV 313
DB 661 EVLGGDGVSGQ---DILEALADLDIELVAT-----LDASORAEIRNYPKTRFTDFVPMHA 713
QY 314 ALSQAQLTIHGKNTVLDALASRTPLALPLAFDOPGVASRIVYHIGKRASRFTTSHA 373
DB 714 LIPSCSAIIHGGAGTATAVINAVPOVMLAELWDAPVKARVAEQAG-----PFLPPAE 769
QY 374 LARQ-IRSLTN-TDYPQRTMKIOAALRLAGGTPAAADIVEQAMR 416
DB 770 LTPQAVRDAVRIILDDPSVATAAHRLEETFGDPTPAGIVPELER 814

RESULT 11

US-08-311-731A-7
Sequence 7, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: MYCOBACTERIUM TUBERCULOSIS

J-08-311-731A-7

Query Match 6.4%; Score 142; DB 4; Length 463;

Best Local Similarity 21.5%; Pred. No. 4.2e-06;

Matches 98; Conservative 69; Mismatches 188; Indels 100; Gaps 22;

20 NLAQELVARGHVFQFQ-----HDCKALVTGSDIGFTVGLQTHPRGSLHLHLAAHP 74

21 NAAEPVARRQRIUVGEAALTAHVRFVLARS-----DPSRYEVHACDP 68

75 -----LGP-----SMRLINEMARTSDMLCRELPAAHALQIE---GVIV 111

69 RFNKLGLPLFPFHPHPTVVPSEVLLKIAQGRLFYNTTRLEKYIAADRKILNFIADPVV 128

112 DQMEPAGAVVAEASGLPVSACAL--PLNREPLPLAVMPFE--YGTSDAARERTTSE 167

129 GENRLSLVSARLAGIPVIAANAYNSPOAR--RPPLDPVPWFVFGVRPV-----SILY 182

168 KIY-----DWLMRRHD--RVIAHACRMGLAPREKLHCHCFSPLAQISQLIPELD 214

183 RLYRPLIFALVCLPLNMLRRKHLSSGLWDLCPRTDGDVLY-----ADVPDLVPTYN 236

215 FPRKALPDCFHVGB-LRQPGGTCSSTSYFSPDKSRIIPASLGTLOGHRYGLPRTIAXA 273

237 -----LPANHYLGVPLWSPDVXKPTWWSLPT-DRPIIYATLGGSGK--NLLQVINA 288

274 CEEVDAQLLHACG--GLSATQAGELARGCPIQVVDPADQSAALSQAQLTTHGMMTVL 331

289 LADLPVTVIATAGNHLKNVPANAF-----VADVLPGEAAARLAVVLCNGSGSPTTQ 341

332 DAIAERTPLALPLAFDPQGVASRIVYHIG--KRASRTTSHALARQISLLTNTDYPQ 389

342 QALAGVFIPLGPNNDQHLNMEALERAGAVLLRTERLNT-BGVAAAVQVLSGAEFR- 399

390 RMTKQAAALRAGTTPAAADIVQAMRQCQVLSG 424

400 -----QAARL-----PKESDQTLPSRSTSKVRG 425

RESULT 12

3-09-356-806-113

Sequence 113, Application US/09356806

Patent No. 6586175

GENERAL INFORMATION:

APPLICANT: Penny, Laura

APPLICANT: Galvin, Margaret

APPLICANT: Miller, Andrew

APPLICANT: Reidy, Michael

TITLE OF INVENTION: Genotyping Human

TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and

FILE REFERENCE: SEQ-22PRV2

CURRENT APPLICATION NUMBER: US/09/356,806

CURRENT FILING DATE: 1999-07-20

NUMBER OF SEQ ID NOS: 164

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 113

LENGTH: 530

TYPE: PRT

ORGANISM: H. sapiens

3-09-356-806-113

Query Match

Best Local Similarity 20.2%; Pred. No. 1.2e-05;

Matches 91; Conservative 75; Mismatches 193; Indels 91; Gaps 18;

6 VIAPFFSHVRLQNLQABLVARHVRVTFQHQDCKALVTGSDIGFTVGLQTHPPGSL 65

27 LVMPTEYSEWINKTIIEBLVQRGHEVTLTS-SASTLVNASK--SSAILEVYPTS--- 80

66 HELHLAAHPLGFSMLRLIN-----EXARTSDMLCRE----- 96

81 ----LTKNDLEDLSLKILDRWIYGVSKNTFWSYFSQLQELCWEYDYSNKLCKDAVLNKK 136
97 LPAAFHALQIGSVIVDQKEPAGAVVAEASGLPFV-----SVACALPLNRE-----PGLPLA 147
137 LAMKLOESKDFVILADANPCGBELAELEFNIPFLYSLSRFSVGYTFEKNGCGFLFPFVSVP 196
148 VMPFEGTSDAARER-----YTTSEKIYDLMRRHDRVIAHACR-----MGLA 191
197 VVMSLSQMFIMERIKNMIMKLYDFDFWFIYD--LKKWDQFYSEVLGRPTTLFETMGKA 254
192 PREKLHCHCFSPLAQISQLIPELDPRKALPDCFHVGPL--RQPGTSGSSTSYFPPSPDK 249
255 ENMLRTYV-----DFEPRPPLPN--VDFVGLHCKPKAPLPEKMEEFVQSSGE 302
250 PRIFA-SLGT-LOGHRYGLFRTIAKACBEVDAQLLHACGGLSATQAGELARGDIOVVD 307
303 NGIVVFSGLSMISNMBESANMIASALAQIPQKVLWRFDGKKPNT-----LGSNTRLYK 356
308 PADQSAALS--QAQLTTHGMMTVLDALASRTPLALPLAFDPQGVASRIVYHIGKRA 365
357 WLPQNDLGHPKTAKFTHGNGTNGIYEALYHIGTIPWGIPLFPADQHDNIAMHAKKAALSV 416
366 S-RFTTSHALARQIRSLTNTDYPQRTXI 394
417 DIRTWSSRDLNALKSVINDPVYKENVMKL 446

RESULT 13

PCT-US92-00282-6

Sequence 6, Application PC/TUS9200282

GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.

TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

TITLE OF INVENTION: THEREIN.

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESS: CUSHMAN DARBY & CUSHMAN

STREET: 1635 L STREET, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20036-5601

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/00282

FILING DATE: 19920110

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 531 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US92-00282-6

Query Match 6.2%; Score 138; DB 5; Length 531;

Best Local Similarity 19.2%; Pred. No. 1.4e-05;

Matches 91; Conservative 72; Mismatches 169; Indels 142; Gaps 20;

6 VIAPFFSHVRLQNALQELVARGHRTVFQOHDCKALVTGSDIGFQTVGLQTHPPGSL 65
29 LVFPWESHWLSDVRELHARGH-----QAVLAPEV-----FVHKGEDF 71
66 HLLHLAHP-----LG-----PSMLRLINEMARTSDMLCRLPAAFH- 102
72 FTLQTYAFFYTKEEYQREILGNKAKGPEQHEVKTFFETWASIKKFFDLVANSACALLN 131
103 -----ALQIEGVIVDQMEPAGVAEASGLPFV-----SVACAL-----PLNRE 141
132 KTLIQQLNSSFDVILDPVFPCCALLAKVLOIPAVEFFLSVPCGIDYEATQCPKSSVI 191
142 PGLPLAVMFFYGTSDAAREYTTSEKIYDMLARRHDRVIAHACRGLAPREKHLHCF 201
192 ENL-LTML-----SDHMTFQRVKMMLYPLTLK-----YICHLSTIPYESL---AS 233
202 PLAQISQILPEL-----DFFRKALPDCHFVAG-----PLRQ----- 232
234 ELLQREMSLVLELHSHASWLVFPGDFVDFYPRIMPNNVFGGNCVKKPLSEFAYVN 293
233 ---PQG-----TPGSSSTSYFSPDKPRIFASLGTQGHRYGLFRTIAKACEEVDAL 285
294 ASGEHGIWVPSLGSWVSEIPEKKAMEIAEALG-----RIPQTLWRY 335
286 CGGISATQAGELARGSDIOVDFADQSAALS--QAQLTITGGGNTVLDALASRTPLAL 343
336 TG-----TRPSNLAK--NTILVWLFQNDLLGHKPARAPITHSNGHYGICNGVPMVM 369
344 PLAPDQPVASRIVHVGIGKASRF--TTSHALAQIRSLTNTDYPQMTKIOA 396
390 PLFGQDMNAKMETRGAGVTLNVLEMTADDEALNTALVNNKSYKENIMELSS 443

RESULT 14
US-08-926-258-2
; Sequence 2, Application US/08926258
; Patent No. 5871983
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Solenberg, Patricia J.
; APPLICANT: Treadway, Patti J.
; TITLE OF INVENTION: Glucosyltransferase Gene gtfE From
; TITLE OF INVENTION: Anycolatopsis orientalis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,258
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: P-10430
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-926-258-2
Query Match 6.2% Score 137.5; DB 2; Length 408;
Best Local Similarity 24.3%; Pred. No. 1e-05;
Matches 109; Conservative 43; Mismatches 201; Indels 95; Gaps 24;
QY 15 VRALQNLAQELVARGHRTVFQOHDCKALVTGSDIGFQTVGLQTHPPGSLHLLHLAHP 74
DB 14 VERLVALAVLRERGAERVMCAPDPCADRLAEVDVPHPLPGASARPSAG-----QAKP 66
QY 75 L-GPSMLRLINEMARTSDMLCRLPAAFH---ALQIEGVIVDQMEPAGAV---VAEASG 126
DB 67 LTADMLRFTTETIATQ---FERIPAAAECAAVTTGLL-----AAAGVRSVAEKL 117
QY 127 LP-FVSACALPLNR---EPGLPLAVMFFYGTSDAAREYTTSEKIYDMLARRHDRV 181
DB 118 IPYFVGFCHSPYVSPYAPPPLGEPAPDGDIDIALWERNQS-----AVERYGEPL 171
QY 192 AHHACRGLAPREKHL-HCPSPLAQISQLIPELDPFRKALPDCHFVAGVPLRQPGT 240
DB 172 NSRRAAIGLPVEVDIFGHGYT-----DHPWMAADP---VLAPL-QPTDLDV 215
QY 241 TSYPFSPDKPRIFASLGTQGH-----RYGLFRTIAKACEEVDALQHLAHCGGLSATQ 293
DB 216 TGAMILPDERPISAELEAFLDAGAPPVVLGFGSLRAPADAA-KVAIEAIRAH----- 266
QY 294 AGE---LARG-GDIQVVD-----PA-----DOSAALSQAQLTITGGGNTVLDALASRTPL 340
DB 267 -GHRVILSRGWADIVLPDREDCEFAIGEVNQVLFRRVAIVHGGAGTTHVATRAGVPQ 325
QY 341 LALPLATDQPVASRIVHVGIGK---ASRFTTSHALAQIRSLTNTDYPQMTKIOA 396
DB 326 ILVQIADQPVYARVAELGVGVAHDGPTTETD-----LSAALTALAPETRVR 378
QY 397 ALRL--AGGTPAAADIVEQAMRTQCPVL 422
DB 379 VAEVTQTDGAADLLFAAVTGNQPAV 406

RESULT 15
US-08-120-053-2
; Sequence 2, Application US/09120053
; Patent No. 5932464
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Solenberg, Patricia J.
; APPLICANT: Treadway, Patti J.
; TITLE OF INVENTION: Glucosyltransferase Gene gtfE From
; TITLE OF INVENTION: Anycolatopsis orientalis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,053
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,258
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: P-10430

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3334

TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 408 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

-09-120-053-2

Query Match 6.2%; Score 137.5; DB 2; Length 408;
 Best Local Similarity 24.3%; Pred. No. le-05;
 Matches 109; Conservative 43; Mismatches 201; Indels 95; Gaps 24;

15 VRALQNLAQELVAGHRVTFQOHDCKALVTGSDIGFTVGLQTHPPGSLSHLHLAAHP 74
 14 VEPLVALAVLRERGAERVCMAPPDCAEDLAEDVDVPHPLGASARPSAG-----QAKP 66

75 L-GPSMLRLINEMARTSMLRELPAAPH---ALQIEGVIVDQMEPAGV-----VAEASG 126
 67 LTAEMLRFTTTIATQ---FERIPAAAGCAAVVTTGLL-----AAAGVRSVAEKLG 117

127 LP-FVSVACALPLNR---EPGLPLAVMPFEYGTG-DAARERYTTSEKIYDMLMRHDEVI 181
 118 IPYFVGFHCPSVPSVYAPPPIPGSPAPDGTDIQALWERNQS-----AYRRYGEPL 171

182 AHHACRMGLAPREKLH-KCFSPQLAQISQLISELPFPRKALPDCPHAVGPLRQPGQTPGSS 240
 172 NSRRAAIGLPPVEDIFGRGT-----DHPWMAADP---VLAPL-QPTDLDAVQ 215

241 TSYFSPDKPRIFASLGLTQGH-----RYGLERTIAKACEEVDQALLAHCGGLSATQ 293
 216 TGAWILDPERPISELEAFLDAGAPPVYLGSGSLRAPADAA-KVAIEAIRAH----- 266

294 AGE---LARG-GDIOVD-----FA---DQSAALSQAQLTIHGHNVTVLDAIASNPL 340
 267 -GERVILSGWADLVLPDDREDCEFAIGEVNQVLFRRVAAVIHGGAGTTHVATRAGVPQ 325

341 LALPLAFDPQGVASRIYVHGIGKR-----ASRFTTSHALAQIRSLLTNTDYPQMTKIQ 396
 326 ILVPGIADQPYAARVAELGVGVAGHDGPTPTFDT-----LSAALFNALAPETRVRAEA 378

397 ALRL--AGGTAAADIVEQAMETCQPV 422
 379 VAETVQTDGAADLLEAAVTGNQPAV 406

Search completed: February 29, 2004, 14:54:58
 > time : 16.0973 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:26:38 ; Search time 50.9584 Seconds
(without alignments)

2389.754 Million cell updates/sec

file: US-09-941-947A-28

irect score: 2231

quence: 1 MSHFAVIAPFFSHVRLQN.....EQAMTCQPVLSGQDYATAL 431

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

tal number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2231	100.0	431	5	Aae22312 Pantoea s
2	2231	100.0	431	6	AAO16018 Pantoea s
3	2231	100.0	431	6	ABP96686 Pantoea s
4	1790	80.2	431	2	AAO7464 Polypepti
5	1790	80.2	431	2	AAW87890 Protein e
6	945	42.4	399	2	AAI13989 Zeaxanthi
7	247	11.1	412	4	AAE61974 S. avermi
8	201	9.0	397	6	ABU11976 M. echino
9	183	8.2	432	6	ABM62888 Photorhab
10	183	8.2	470	6	ABP57697 Saccharop
11	181.5	8.1	540	4	ABE64912 Drosophil
12	178.5	8.0	492	4	ABE84216 Amino aci
13	174.5	7.8	541	6	ABU89759 Protein d
14	174.5	7.8	541	7	ADD48913 Human pro
15	171.5	7.7	414	2	AAE93982 Saccharot
16	170	7.6	527	6	ABU54566 Human NOV
17	167	7.5	422	3	AAV83786 S. fredia
18	167	7.5	527	6	AAU29284 Human PRO
19	167	7.5	527	5	AAU77927 Human dru
20	167	7.5	527	5	ABE71237 Human 326
21	167	7.5	527	5	AAE15434 Human dru
22	167	7.5	527	6	ABU58660 Human PRO
23	167	7.5	527	6	ABU88208 Novel hum
24	167	7.5	527	6	ABU84523 Human sec
25	167	7.5	527	6	ABE66397 Human sec

26	167	7.5	527	6	ABR65787 Human sec
27	167	7.5	527	6	ABU99727 Human sec
28	167	7.5	527	6	ABU82966 Human PRO
29	167	7.5	527	6	ABU90087 Novel hum
30	167	7.5	527	6	ABR68336 Human sec
31	167	7.5	527	6	ABU96389 Novel hum
32	167	7.5	527	6	ABU92820 Human sec
33	167	7.5	527	6	ABO08897 Human sec
34	167	7.5	527	6	ABO02949 Human sec
35	167	7.5	527	6	ABR75103 Human sec
36	167	7.5	527	6	ABR94865 Human sec
37	167	7.5	527	6	ABU85838 Human PRO
38	167	7.5	527	6	ABU98998 Novel hum
39	167	7.5	527	6	ABU98213 Novel hum
40	167	7.5	527	6	ABU91919 Novel hum
41	167	7.5	527	6	ABU89612 Human PRO
42	167	7.5	527	6	ABU86453 Human sec
43	167	7.5	527	6	ABU67666 Human sec
44	167	7.5	527	6	ABU80694 Human PRO
45	167	7.5	527	6	ABR99612 Human sec

ALIGNMENTS

RESULT 1

AAE22312
ID AAE22312 standard; protein; 431 AA.

XX AAE22312;

XX 25-JUL-2002 (first entry)

DT Pantoea stewartii zeaxanthin glucosyl transferase (CrtX) enzyme.

DE Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; zeaxanthin glucosyl transferase; CrtX.

XX Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229858P.

XX 01-SEP-2000; 2000US-0229907P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odom JM, Picataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35510.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates.

XX Claim 25; Page 136-137; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The

XX method comprises a transformed metabolising host cell, comprising

XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX control of regulatory sequences, and contacting the host cell with carbon

XX substrate to produce a carotenoid compound. The method is useful for

XX producing carotenoid compounds such as antheraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

C the carotenoid biosynthetic pathway and which metabolise single carbon
C substrates. The carotenoids have potent anti-oxidant properties useful in
C diet, and aquaculture elements. The carotenoids are also useful as
C intermediates in the synthesis of steroids flavours and fragrances and
C compounds for potential electro-optic applications. The present sequence
C is Pantoea stewartii zeaxanthin glucosyl transferase (CrtX) enzyme used
C in the invention

X
Q Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.4e-215;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 MSHFAVIAPPPFVSHVRLALQNLAEQELVARGHRTVFFQOHDCKALVTGSDIGFOTVGLQTHP 60
b 1 MSHFAVIAPPPFVSHVRLALQNLAEQELVARGHRTVFFQOHDCKALVTGSDIGFOTVGLQTHP 60
Y 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
b 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
Y 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFFYGTSDAARERTTSEKIYDMLMRHDRV 180
b 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFFYGTSDAARERTTSEKIYDMLMRHDRV 180
Y 181 IAHACRMGLAPREKLUHCHFSPLAQISQLIPELDFPRKALPDCFHAGVPLRPOQGTGSS 240
b 181 IAHACRMGLAPREKLUHCHFSPLAQISQLIPELDFPRKALPDCFHAGVPLRPOQGTGSS 240
Y 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
b 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Y 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
b 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
Y 361 IGKASRFTTSHALAQIRSLTNTDYPQRTKIQAALRLAGTTPAAADIVEQAMTCOP 420
b 361 IGKASRFTTSHALAQIRSLTNTDYPQRTKIQAALRLAGTTPAAADIVEQAMTCOP 420
Y 421 VLSQDYATATL 431
b 421 VLSQDYATATL 431

RESULT 2

AA016018 standard; protein; 431 AA.

AA016018;

20-FEB-2003 (first entry)

Pantoea stewartii zeaxanthin pyrophosphate synthase.

Carotenoid; crt.

Pantoea stewartii.

WO200279395-A2.

10-OCT-2002.

25-JAN-2002; 2002WO-US002124.

26-JAN-2001; 2001US-0264329P.

04-MAY-2001; 2001US-0286984P.

(CRGI) CARGILL INC.

De Souza ML, Kollmann SR, May CA, Schroeder WA;

XX WPI; 2003-075455/07.
DR N-PSDB; ABT14190.

XX Novel isolated nucleic acid useful e.g. to engineer host cells with the
PT ability to produce particular carotenoids and polypeptides useful in cell
PT -free systems to make particular carotenoids.

XX Claim 7; Page 58-59; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
CC carotenoid (crt)-related proteins. The crt-related DNA and protein
CC sequences of the invention are useful for engineering cells which are
CC able to produce carotenoids. The present amino acid sequence represents a
CC crt-related protein of the invention

XX Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 6; Length 431;

Best Local Similarity 100.0%; Pred. No. 4.4e-215;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSHFAVIAPPPFVSHVRLALQNLAEQELVARGHRTVFFQOHDCKALVTGSDIGFOTVGLQTHP 60

DB 1 MSHFAVIAPPPFVSHVRLALQNLAEQELVARGHRTVFFQOHDCKALVTGSDIGFOTVGLQTHP 60

QY 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120

DB 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120

QY 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFFYGTSDAARERTTSEKIYDMLMRHDRV 180

DB 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFFYGTSDAARERTTSEKIYDMLMRHDRV 180

QY 181 IAHACRMGLAPREKLUHCHFSPLAQISQLIPELDFPRKALPDCFHAGVPLRPOQGTGSS 240

DB 181 IAHACRMGLAPREKLUHCHFSPLAQISQLIPELDFPRKALPDCFHAGVPLRPOQGTGSS 240

QY 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300

DB 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300

QY 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360

DB 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360

QY 361 IGKASRFTTSHALAQIRSLTNTDYPQRTKIQAALRLAGTTPAAADIVEQAMTCOP 420

DB 361 IGKASRFTTSHALAQIRSLTNTDYPQRTKIQAALRLAGTTPAAADIVEQAMTCOP 420

QY 421 VLSQDYATATL 431

DB 421 VLSQDYATATL 431

RESULT 3

ABP96686 standard; protein; 431 AA.

ABP96686;

03-JUN-2003 (first entry)

Pantoea stewartii zeaxanthin glucosyl transferase SEQ ID NO:4.

Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;

crtI; crtB; crtZ; zeaxanthin glucosyl transferase; enzyme; phytoene;

carotenoid.

Pantoea stewartii.

WO2003016503-A2.

XX

27-FEB-2003.
 15-AUG-2002; 2002WO-US026647.
 15-AUG-2001; 2001US-0312646P.
 (DUFO) DU PONT DE NEMOURS & CO E I.
 Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PB,
 WPI; 2003-269323/26.
 N-PSDB; ACC44760.
 Novel nucleic acid molecule isolated from *Pantoea stewartii* encoding a carotenoid biosynthetic enzyme, useful for regulating carotenoid biosynthesis in an organism.
 Claim 4; Page 59-60; 68pp; English.
 The present invention describes *Pantoea stewartii* carotenoid biosynthetic enzymes (I). More specifically described are the geranylgeranyl pyrophosphate synthase (crtB), zeaxanthin glucosyl transferase (crtX), lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating carotenoid biosynthesis in an organism, by over-expressing (I) in an organism, such that the carotenoid biosynthesis is altered in the organism. (II) and the genes encoding (I) are useful for converting phytoene to the carotenoids, for creating recombinant organisms that have the ability to produce various carotenoid compounds, and also for enhancing or manipulating carotenoid compounds. (I) can also be used for producing gene products having enhanced or altered activity

Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 6; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.4e-215;
 Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSHFAVIAPPPSHVRALQNLQALVARGHRTVFQQHDCALVTGSDIGFOTVGLQTHP 60
 1 MSHFAVIAPPPSHVRALQNLQALVARGHRTVFQQHDCALVTGSDIGFOTVGLQTHP 60
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 121 VAASGLPTVSVACALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 121 VAASGLPTVSVACALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 181 IAHCRCWGLAPREKHLHCFSPLAQISQLIPELDFPRKALPDCFHAVGLPQPGTQSS 240
 181 IAHCRCWGLAPREKHLHCFSPLAQISQLIPELDFPRKALPDCFHAVGLPQPGTQSS 240
 241 TSFFPDPKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLHACGSLATAGELARG 300
 241 TSFFPDPKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLHACGSLATAGELARG 300
 301 GDIVVDQADQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 301 GDIVVDQADQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 361 IGRASRFTTSHALAQIRSLTNTDYPORMTKIOAALRLAGGTAAADIIVEQAMRTQCP 420
 361 IGRASRFTTSHALAQIRSLTNTDYPORMTKIOAALRLAGGTAAADIIVEQAMRTQCP 420
 421 VLSGGDYATAL 431
 421 VLSGGDYATAL 431

RESULT 4

AAR07464
 ID AAR07464 standard; protein; 431 AA.
 AC XX
 AC AAR07464;
 XX XX
 DT 24-OCT-2003 (revised)
 DT 28-JAN-1991 (first entry)
 XX XX
 DE Polypeptide with enzymatic activity for the conversion of zeaxanthin into zeaxanthin diglucoside.
 DE DE
 XX Carotenoid biosynthesis; vitamin A; cancer; food coloring.
 XX OS
 XX *Pantoea ananatis*.
 XX PN EP933690-A.
 XX XX
 PD 24-OCT-1990.
 XX XX
 PF 20-APR-1990; 90EP-00107493.
 XX PR 21-APR-1989; 89JP-00103078.
 PR 05-MAR-1990; 90JP-00053255.
 XX XX
 PA (KIRI) KIRIN BEER KK.
 XX XX
 PI Misawa N, Kobayashi K, Nakamura K;
 XX WPI; 1990-322212/43.
 DR N-PSDB; AAQ06294.
 DR XX
 PT DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.
 PT XX
 PS Claim 2; Fig 2; 40pp; English.
 XX XX
 CC Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)
 CC CC
 XX Sequence 431 AA;

Query Match 80.2%; Score 1790; DB 2; Length 431;
 Best Local Similarity 80.5%; Pred. No. 9.2e-171;
 Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;

1 MSHFAVIAPPPSHVRALQNLQALVARGHRTVFQQHDCALVTGSDIGFOTVGLQTHP 60
 1 MSHFAVIAPPPSHVRALQNLQALVARGHRTVFQQHDCALVTGSDIGFOTVGLQTHP 60
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 121 VAASGLPTVSVACALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 121 VAASGLPTVSVACALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 181 IAHCRCWGLAPREKHLHCFSPLAQISQLIPELDFPRKALPDCFHAVGLPQPGTQSS 240
 181 IAHCRCWGLAPREKHLHCFSPLAQISQLIPELDFPRKALPDCFHAVGLPQPGTQSS 240
 241 TSFFPDPKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLHACGSLATAGELARG 300
 241 TSFFPDPKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLHACGSLATAGELARG 300
 301 GDIVVDQADQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 301 GDIVVDQADQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 361 IGRASRFTTSHALAQIRSLTNTDYPORMTKIOAALRLAGGTAAADIIVEQAMRTQCP 420
 361 IGRASRFTTSHALAQIRSLTNTDYPORMTKIOAALRLAGGTAAADIIVEQAMRTQCP 420


```

Sequence 399 AA;
Query Match 42.4%; Score 945; DB 2; Length 399;
Best Local Similarity 47.9%; Pred. No. 6.9e-86;
Matches 201; Conservative 59; Mismatches 126; Indels 34; Gaps 7;

1 MSHFAVAPPFFSHVRALQNLQALVARGHRTVFFQCHDCKALVTGSDIGF---QTVGLQ 57
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSHEAVAPPFLYSHAVAVHALALQMAQRGHRVTF-----LTGNVASLARQETERVA 51

58 THP-PGSLSHLLHLAAHPLGSLMLRLINEMARTSDMLCRELPRAFHALQIEGVVDOME 116
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 FYLPASVQQQQRNVQQQSGNGLRLTAAMSSLDVLCQLPAILQRLAVDALVIDMEP 111
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

117 AGAVVASAGLPLFVSACALPLNREPGLPLAVMPFEYGTSDAARERTTSEKIYDMLMR 176
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 AGSLVAEALGLPFISIALCPVNE--LPLVMPFFHYAEDKRARPOVSEIRIYDALMYP 169
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

177 HORVIAHACRGLAPREKLHCHCSPLAQISOLIPELDPRKALPDGFHAGVLPQGT 236
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 DGQITLREARQFGLPERRRDECLSPAQISQSVPALDFPRALPNCFTTVGALR 224
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 PGSSTVPPSP-----DKPRIFASLTQGHRYGLFRTIAKACEEVDQAQLLAHCGGL 289
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 -----YQPPPOVERSPRSTRIIFASLTQGHRLFLFKIARACASVGAETLAHCDGL 278
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

250 SATQAGELARGDITQVDFADQSAALSOAQLTTHCGMNTVLDIAISRTPLIALPLAFDQ 349
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 TPAQADSLYCGA-TEVVSFVDPQRYVAEAKLIVTHGGINTVLDALAAATPVLAVPLSFDQ 337
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

350 PGVASRIYTHGIGKASRFTTSHALARCIRSLTNTDYPQRMTKIQALRLAGTAAAD 409
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 PAVARLVINGLRVSRFARQETLADEIAQLLGDTELHERVATPAQQLNDAGTFRGCD 397
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 7
ABU1974
AAB61974 standard; protein; 412 AA.
AAB61974;
14-MAY-2001 (first entry)
S. avermitilis ORF2 (AverB) polypeptide.
Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;
oleandrose; ORF2; AverB; glycosyl transferase.
Streptomyces avermitilis.
WO200109155-A1.
08-FEB-2001.
26-JUL-2000; 2000WO-US020331.
30-JUL-1999; 99US-0146699P.
(MERI) MERCK & CO INC.
Macneil DJ, Occi J, Gewain KM;
WPI; 2001-182929/18.
N-PSDB; AAC85191.
Novel nucleic acid fragments of Streptomyces avermitilis genome useful
for enzymatic, biochemical, biosynthetic and diagnostic purposes.
Disclosure; Page 20; 63pp; English.
The invention relates to a 10 kb genomic DNA isolated from S. avermitilis
that contains genes encoding proteins for glycosylation of avermectin

```

```

CC aglycones. The polypeptides can be expressed by standard recombinant
CC methodology. The gene and the encoded polypeptides together with other
CC enzymatically active polypeptides, are useful to perform combinatorial
CC biocatalysis in vitro and in vivo in a host cell. They are useful for
CC performing biotransformations on macrolide compounds including avermectin
CC or other macrolide aglycones. The sequences are also useful in vivo in a
CC bacterial host, in vitro in combination with an actinomycete
CC fermentation, and in vitro in combination with enzymatically active
CC polypeptides that are not from the avermectin biosynthetic pathway to
CC effect the synthesis of a pharmaceutically active compound, e.g. an
CC antibiotic. Sequences AAB61973-981 represent polypeptides in the
CC avermectin biosynthetic pathway, involved in the synthesis and/or
CC addition of oleandrose to avermectin aglycones. The present sequence is
CC the S. avermitilis ORF2 (AverB) polypeptide having the activity of
CC glycosyl transferase
XX
SQ Sequence 412 AA;
Query Match 11.1%; Score 247; DB 4; Length 412;
Best Local Similarity 21.3%; Pred. No. 1e-15;
Matches 102; Conservative 76; Mismatches 162; Indels 138; Gaps 15;

QY 3 HFVATAPPFFSHVRALQNLQALVARGHRTVFFQCHDCKALVTGSDI----- 49
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 4 HFLFMSAPFWGHVFPFLAVAEELVHRGHVTF-----VTGAEMADAVRSVGADFLR 54
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 50 ---GFQTGLQTHPPGSLSHLLHLAAHPLGSLMLRLINEMA--RTSDMLCRELPAAFAHAL 104
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 55 YESAFEGVDYMLRTEAPNAIPMLYDEGMSMLRSVEEHVGKDVPLVDIATSLN-- 112
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 105 QIEGVVDOMEPAAGVAAEAGLFPFVSACALPLNREFGLPLAVMPFEYGTSDAARBYT 164
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 113 -----VGRVLAASMSRPAMTV-----IPLF-----ASNGRFS 139
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 165 TSEKIYDMLMRHURVIAHACRAGLAPREKLHCHCS---PLAOSLIPEL----- 213
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 TMSQSVLD-----PDSAQVSAPPFR-----FSOMELFGALVPLALLLVSRG 183
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 214 -----DPPRKALPDCHF-----AVGPLRQPGTGSTSTVFPSPDK 249
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 ITEPVDDFLSGPEDFNLCVLPFAQYAGDTDFERFAPFVPGCLGKRGIGWTP--PGSGH 241
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 250 PRIFASLTQGHRYGLFRTIAKACEEVDQAQLLAHCGGLSATQAGELARGD----- 302
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 242 PVLISLTGTVFNRLSFFRTFVRAFTDVPVHVVIS-----LGKGVDPDVLRLPL 289
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 303 ---IQVDFADQSAALSOAQLTTHCGMNTVLDIAISRTPLIALPLAFDQPGVASRIYTH 359
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 290 PENVEVHRVPHAVLEHARALVTHGGTGVMEALHAGCPVLVMPFSRDAQVTGRIAEAL 349
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 360 GIGKRAS-RFTTSHALARQIRSLTNTDYPQRMTKIQALRLAGTAAADIVQAMR 416
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 350 GLGRMVQPEEVTATTLRRHVLDIISDDAITQVRQMRATVEAGGALRAADETERFLR 407
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 8
ABU1976
ID ABU1976 standard; protein; 397 AA.
XX
AC ABU1976;
XX
DT 23-OCT-2003 (revised)
DT 13-FEB-2003 (first entry)
XX
DE M. echinospora calicheamycin biosynthesis protein CalN.
XX
KW Calicheamycin biosynthetic gene cluster; aryltetrasaccharide; aglycone;
KW calicheamycin resistance; nonchromoprotein enediynes; enediynes resistance;
KW bone marrow cell; enzyme.
XX
OS Micromonospora echinospora; spp. calichensis.
XX
WO200279465-A2.

```


NT New isolated nucleic acid detection reagent for detecting 1000 or more
 TT genes from Drosophila and for elucidating cell signaling and cell-cell
 TX interactions.

XX Disclosure: SEQ ID NO 21528; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signaling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 XX sequences (ABU161840-ABU16175) and the encoded proteins (ABU57737-
 XX ABU572072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 540 AA;

XX Query Match 8.1%; Score 181.5; DB 4; Length 540;
 XX Best Local Similarity 23.7%; Pred. No. 5.8e-09;
 XX Matches 109; Conservative 66; Mismatches 202; Indels 83; Gaps 19;

Y 7 IAPFFSHVRAQLONLACELVARGHVTFFQHQCKALVTG-----SD 48
 b 36 IPSPF-----QKVRFLIKALVERGHKVTWTPADYPAKIDGVRHVRPMLNQLMKNMD 91
 Y 49 IGFVTGLQTHPPGSLHLLHAHPGLPSMLRLINEMARTSMCLRELPAAPHALQIEG 108
 b 92 QFFDALGDKWRREGVLVSTIFNVSHAI-----LNNNGVQ---MLMRDKSIRFDMIMVEA 142
 Y 109 VIVDQMEPAGAVVAASGLFPVSACA-----LPLNRBPGLPVMFEFGTSDAAR 160
 b 143 SHLDALYG-----LAEFNATLLGISEMHTWHDYLAGNLAPSVYPIGPNFALDNTFL 198
 Y 161 ER-----YTTSEKIYDMLMRHRDRIAHACRMGLAPREKLRHCFSPLAQISOLIPELDP 215
 b 199 SRWNWIIYIEKLERLVRPAQVR--FK--KFGYPAEKIDEL---RAEFSVILVNSHF 253
 Y 216 P-----RKALPDCFHAVG--PIRQPGTGGSS--TSYFSPDPKPRIFASLTGLQGHRYGLPT 269
 b 254 SMGRVRANVPNIIEVAGVHLSEPPPCGABQLQYLDIAERGAIFYSMGQDILIKY----- 308
 Y 270 IAKACEEVDQAOLLAHCGGLSATQ-----AGELA-----RGDIOQVDFADQSAALS--OAQ 319
 b 309 ---LPENMKQLLVF---LQMKQVITWKSLSMLANKSENIYMDKVPQWVLAHPNLR 362
 Y 320 LTITHGGMNTVDAASRTPLALPLAPDQGVASRIVYHIGKGRAS--RTTSHALAROI 378
 b 363 LFITHGGLQSWAIDNGVPMGLPLFDQFNHIVQLAGMAKVLDPNLDNADTLIETI 422
 Y 379 RSLLTNTDYPQRTKIQALRLAGGTAAADI--VEQAMR 416
 b 423 KELLNPSYAORAKEMASPRDRPMSPLDTAIWTTETALR 462

RESULT 12

AB84216

ID AB84216 standard; protein; 492 AA.

AC AB84216;

XX 06-AUG-2001 (first entry)

XX Amino acid sequence of an UDP-glucose:aglycon-glucosyltransferase.

XX UDP-glucose:aglycon-glucosyltransferase; UDP-GAG; cyanohydrin; terpenoid;
 KW glucose; transgenic plant; cyanogenic glucoside biosynthesis;
 KW pathogen resistance; herbivore response.

XX Sorghum bicolor.

XX WO200140491-A2.

XX 07-JUN-2001.
 XX 29-NOV-2000; 2000WO-EP011982.
 XX 01-DEC-1999; 99EP-00123838.
 XX (LUMI-) LUMINIS PTY LTD.
 XX (UVRO-) UNIV ROYAL VETERINARY & AGRIC.
 XX Hoef P, Moeller BL, Jones PR;
 XX WPI; 2001-374846/39.
 XX N-PSDB; AAF90242.
 XX DNA molecule coding for UDP-glucose:aglycon-glucosyltransferase
 XX conjugating cyanohydrin, terpenoid or phenyl derivative to glucose, for
 XX producing transgenic plants having modified cyanogenic glucoside
 XX biosynthesis.
 XX Claim 7; Page 27-29; 31pp; English.
 XX The present sequence represents a UDP-glucose:aglycon-glucosyltransferase
 XX (UDP-GAG) polypeptide. The enzyme conjugates a cyanohydrin, terpenoid,
 XX phenyl derivative or hexanolderivative to glucose. UDP-GAG polynucleotides
 XX are useful for producing transgenic plants having modified cyanogenic
 XX glucoside biosynthesis. Constitutive, inducible or tissue-specific
 XX expression of UDP-GAG is useful for obtaining transgenic cyanogenic
 XX plants with altered resistance to pathogens and herbivore responses
 XX Sequence 492 AA;

XX Query Match 8.0%; Score 178.5; DB 4; Length 492;
 XX Best Local Similarity 23.2%; Pred. No. 1e-08;
 XX Matches 113; Conservative 73; Mismatches 191; Indels 111; Gaps 22;
 QY 3 HPAVIAPFTSHVRAQLONLACELVARGHVTFFQHQCKALVTGSDIGPQTGLQTHPP 61
 Db 12 HWLVPPFGQGHVAPLQMLARLLHARGARVTFYTYTNYRLLRAK--GEAAV-----RPP 65
 QY 62 GSLSHLLHAHPGLPSMLRLINEMARTSMCLRELPAAPFHAL-----QIEG----- 108
 Db 66 ATSSARFRIEVIDDGLSLVQNDVGLVDSLRKNCCLHPFRALLRLRGQSEVGODAPPVT 125
 QY 109 -VIVDQMEPAGAVVAASGLP-----FVSVACAL--PLNRBPGLPVMFEFGTSDAARE 161
 Db 126 CWGDDVWVTFAAAAAREAGIPEVQFTASACGLLGLYHGYELVERGLVPER-DASLLADD 184
 QY 162 RYTTSEKIYDML-----MRHRDRIAHACK-----MGLAPREKLRHCFSPLAQISOL 209
 Db 185 DYLDTP--LEWVPGKSHMLDMDPT---FCRTTDPDDVYSATLQOMESAAGSKALILNT 239
 QY 210 IPELDFPRKALPDC-----FHAVGELRQPGTGGSTSYFSPD----- 248
 Db 240 LYLEL---KQVVDALAAFFPPIYTVGFLAEVIASSDSASAGLAANDISIWQEDTRCLSWL 296
 QY 249 --KPR---IPASLTGLQGHRYGLFTTAKACEEVDQAOLLAHCGG----- 288
 Db 297 DGKSGSVVYVNFSGM-----AVMTAAQAREFALGLASCSPLWVKRPDQVVEGEE 347
 QY 289 --LSATQAGELARGDIIQVDFADQSAALSQAQ--LTITHGGMNTVDAASRTPLIALP 344
 Db 348 VLLPEALLDVARGGL-VVFWCPQAAVLKHAAGVLFVSHCGWNSLLEATAAGQPVILAWP 406
 QY 345 LAFDQGVASRIV-YHIGKGRASFTTSHALAROIRSLLTNTDYPQRTK-----IOA 396
 Db 407 CHGEQTNCRQLCEVWNGAQLPREVESGAVRLVREKVMVGDIGKEKRAAEWKAAREA 466
 QY 397 ALRLAGGT 404
 Db 467 AARKGAS 474

RESULT 13
ABU89759
ABU89759 standard; protein; 541 AA.
ABU89759;
10-JUL-2003 (first entry)
Protein differentially expressed in cardiovascular disease #53.
Cardiovascular disease; arteriosclerosis; ischaemia; angina pectoris;
myocardial infarction; cardiast; antiarteriosclerotic; antianginal;
gene therapy; differential gene expression.
Homo sapiens.
WO2003031650-A2.
17-APR-2003.
02-OCT-2002; 2002WO-EP011034.
08-OCT-2001; 2001GB-00024145.
(FARB) BAYER AG.
Munnes M, Gehrman M, Wick M, Schmitz G;
WPI; 2003-403108/38.
N-PSDB; ACAB9932.
Predicting, diagnosing or prognosing a cardiovascular disease, e.g.
angina, ischemia, myocardial infarction or arteriosclerosis by detection
of a polynucleotide in a biological sample comprises detecting a
hybridization complex.
Claim 3; Page 406-408; 454pp; English.
The invention describes a method of predicting, diagnosing or prognosing
a cardiovascular disease by detection of a polynucleotide in a biological
sample comprises hybridising at least one of the polynucleotide to a
nucleic acid material of a biological sample, thus forming a
hybridisation complex, and detecting the hybridisation complex. The
polynucleotides, polypeptides, antisense molecule, antibody and reagent
are useful for preparing compositions for preventing, predicting or
diagnosing, or a medicament for treating a cardiovascular disease, e.g.
arteriosclerosis, ischaemia, angina pectoris, or myocardial infarction.
This sequence represents a protein identified in the invention a being
differentially expressed in individuals with cardiovascular disease
Sequence 541 AA;
Query Match 7.8%; Score 174.5; DB 6; Length 541;
Best Local Similarity 20.1%; Pred. No. 2.9e-08;
Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;
6 VIAPP--PFSHVRALONLAQELVARGHRTVFFQHDCKALVTGSDIG--FQTVGLQTHP-- 60
24 IIVPTIMPESHMYIFKTLASALHGHGHTVFL-----LSEGRDIAPSNHYSLQRYPGI 76
61 -----PGSLSHLLHAAHPLGSMRLINEMARTSDMLCRELPAAFH 102
77 FNSITSDAFLQSMENITSGRLTAI-----ELFDILDHYTKCDLM-----VGNH 121
103 ALQIEG-----VIVDMEPAGAVABASGLPFVSVACALPLNRSPG--LFLAVMPFE 152
122 AL-IQLKKEKEDLLVDPNDMCGFVAHLLGVKAVESTGLWYPAEYVADAPLAYVPE-E 179
153 YGTSDAAREYTTSEKIDWLMRHRDVIHAHACWGLA-----PREKLHCFSPLAIOISQ 208
180 FNSLUTDMN-----LLQRMKNTGVLIISKLGVSLVLPK-----FERIMOKYN 223
209 LIPE-----LDFPRKALPDCFHAVGPLRQPGTQPGSGSTSYFFSPD 248

Db 224 LLEPEKSMYDLVHGSSLMMLCTDVALEFPFPLPNVYVGGILTKPAS-----PLPE 274
Qy 249 KPIFASLGTLOGHRYGLFRTIAKACEVDQAOLLAHCGGUSATOAGELAR----- 299
Db 275 DIQRWVNGANEHGFVLVSFGAGVYLSSEDIANKL-----AGALGLPQKVIWRF 323
Qy 300 -----GGIOVVDFADOSAALSOAQLT--ITHGGMNTVLDALASRTPLLALPLAFDQ 349
Db 324 SGPKPKNUNNTKLEWLPQNDLGHSKIKAFLSHGGLNSIFETMYHGVVGVPLFGDH 383
Qy 350 PGVASRIYVHIGIGKAS-RFTTSHALAQIRSLTNTDYPQRMTKIQALRLAGTTP 405
Db 384 YDTMTRVQAKGKGLLEKTKVTEKELYEALVKVINNPSYRQRAQKLSIHKDQPGHP 440

RESULT 14

ADD48913
ID ADD48913 standard; protein; 541 AA.

XX
AC ADD48913;

XX
DT 29-JAN-2004 (first entry)

XX
DE Human Protein Q16880, SEQ ID NO 14624.

XX
KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX
OS Homo sapiens.

XX
PN WO2003016475-A2.

XX
PD 27-FEB-2003.

XX
PF 14-AUG-2002; 2002WO-US025765.

XX
PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX
PA (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX
PI Woolf C, D'urso D, Befort K, Costigan M;

XX
WPI; 2003-268312/26.

XX
GENRANK; Q16880.

PT New composition comprising two or more isolated polypeptides, useful for
preparing a medicament for treating pain in an animal.

XX
Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
derivative or allelic variation of the nucleic acid sequence. Also
claimed are a vector comprising the novel polynucleotide, a host cell
comprising the vector, a method for identifying a nucleotide sequence
which is differentially regulated in an animal subjected to pain and a
kit to perform the method, an array, a method for identifying an agent
that increases or decreases the expression of the polynucleotide sequence
that is differentially expressed in neuronal tissue of a first animal
subjected to pain, a method for identifying a compound which regulates
the expression of a polynucleotide sequence which is differentially
expressed in an animal subjected to pain, a method for identifying a
compound that regulates the activity of one or more of the
polynucleotides, a method for producing a pharmaceutical composition, a
method for identifying a compound or small molecule that regulates the
activity in an animal of one or more of the polypeptides given in the
specification, a method for identifying a compound useful in treating
pain and a pharmaceutical composition comprising the one or more

polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 541 AA;

Query Match 7.8%; Score 174.5; DB 7; Length 541;
 Best Local Similarity 20.18; Pred. No. 2.9e-08;
 Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;
 6 VIAPP--FFSHVRLQALQELVARGHRVTFQCHDCKALVTSGDIG-FQTGVLQTHP-- 60
 24 IIVPPIMFESHMYIFKTLASALHGRHTVFL-----LSEGRDIAPNSHYSLOQYPI 76
 61 -----PGSLSHLLHAAHPLGSMRLINEMARTSDMLCRELPAAFH 102
 77 FNSTSDAFLOSQRNIFSGRLTAI-----ELFDILDHYTKNCULM-----VGNH 121
 103 ALQIEG-----VVDQMEPAGAVARASGPFVSVACALPLNREPQ--LPLAVMPFE 152
 122 AL-IOGLKKEKFDLLVDPNDCGFIHLLGVKIAVSTGLWPAEYVGAPAPLAYVP-E 179
 153 YGTSDAARERTTSEKIYDMLRRHRDRIAHACRMGLA---PREKLHCPSPLAQISQ 208
 180 FNSLLTDRNN-----LQKQNTGYVLSRIGVSFLVLPK-----YERINQKN 223
 209 LIPE-----LDPRKALPDCFHAGVPLQPCQGTGSGTSYSPSPD 248
 224 LLPEKMYDLVHGSSLMCLTDVALEFPRTLENVYVGGILTKPAS-----PLPE 274
 249 KRFIFASLTQGHRYGLFRITIAKACEVDQAQLLAHCGGLSATQAGELAR----- 299
 275 DLQRNVNGANEHGFVNSFGAGVKYLSIEDIANKL-----ACALGRLPQKVINRF 323
 300 -----GGDIQVVDFAQSAALSQAQLT--ITHGQMTVLDIAISRTPLLALPLAFDQ 349
 324 SGKPKNLGNNTKLEIWMPLQNDLLGHKIKAFLSHGGLNSIFETWVHGVPVWGIPLFGDH 383
 350 PQVASRIVYHIGIKRAS-RFTTSHALARQIRSLTNTDYDQRTKIOAALSLAGGTF 405
 384 YDTWTRVQAKGNGILLEWKVTYKELYEALVKVINNPSYRQAKUSEIHKQDFGHP 440

RESULT 15

AAR93982 standard; protein; 414 AA.

AAR93982;

16-OCT-2003 (revised)

15-AUG-1996 (first entry)

Saccharothrix aerocolonigenes glycosyl transferase.

Glycosyl transferase; glycosylating; indolopyrrolocarbazole;
 Actinomycetales; cancer treatment; anti-tumour.

Lechevalieria aerocolonigenes; (ATCC 39243).

W09534653-A1.

21-DEC-1995.

31-MAY-1995; 95WO-JP001065.

13-JUN-1994; 94JP-00154127.

(BANY) BANYU PHARM CO LTD.

Kojiri K, Suzuki H, Kondo H, Suda H;

WPI; 1996-049691/05.

N-PSDB; AAT29923.

Gene coding for glycosyl transferase - useful in glycosylation of indolo:pyrrolo carbazole derivs.

Claim 4; Page 4-6; 45pp; Japanese.

AAR93982 is a glycosyl transferase enzyme derived from Saccharothrix aerocolonigenes (ATCC 39243). The glycosyl transferase (GT) may be from any organism of the order Actinomycetales. GT is used in a method of glycosylating indolopyrrolocarbazole derivs. Glycosylated indolopyrrolocarbazole derivs. are useful as anti-tumour agents. GTs can be made by recombinant means using new GT-producing strains Streptomyces lividans TK21 (pNGT207) FERM BP-5091 and S. morbariensis BA 13793 (pNGT207) FERM BP-5090. (Updated on 16-OCT-2003 to standardise OS field)

Sequence 414 AA;

Query Match 7.7%; Score 171.5; DB 2; Length 414;
 Best Local Similarity 21.48; Pred. No. 4e-08;
 Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
 6 VIAPPPFSHVRLQALQELVARGHRVTF----- 34
 7 VATTFGDGHNPVPAQEMVSRGHEVRYTGTAKFRSTVERTGARHEPMDADHDFGMPR 66
 35 ---FQCHDCKALVTSGDIGFQTGVLQTHPQCSLSHLHLAAHPLGFSMLRLINEMARTSD 91
 67 EEAFFQHLNGLTGHTGNIAGFRDPIE-----PAADQMTDLL----- 102
 92 MLCRELPAAPHALQIEGVIVDQMEPAGAVARASGLPFFSVACAL-----PLNREP 142
 103 ALLEDFFA-----DVLVTDETFFGAGFVSERTGIFVAMIATSIYVFSRDRTPAGL-- 153
 143 GLPLAVMPFEVGTSDAARERTTSEKIYDML---MERHDRVIAHACRMGLAPREKLHH 198
 154 GLPFS-----SRLGLRLNTVLKQLTDRVMDLERHADVVD---RVGL-PRIR-KG 201
 199 CFSPLAQISQL-----IPELDFPRKALPDCFHAGVPLRQPCGTGSGTSYSPSPD--KPR 251
 202 AFENIMRTFDLYLGTVFSPEYPRGDMPEVRFVFPFVSPAPDPFTPPAWMGELDSGRPV 261
 252 IFASLTQGHRYGLFRITIAKACEVDQAQLLAHCGGLSATQAGELARGG----- 301
 262 VEVTTG-----TVANDAE-----RLLLPAIRALAAEDVLVATTGAPLELEMPA 306
 302 DIQVVDFAQSAALSQAQLTITHGQMTVLDIAISRTPLLALPLAFDQPGQVASRIVYHGI 361
 307 NVRVRFIPFHEALLPHVDAMVTNGYGGVNTALAGVPLV-VARTEKEHYAARVSWSGA 365
 362 GKR-ASRFTTSHALARQIRSLTNTDYDQ 389
 366 GVHLKKRELSERDIRRAVAVL---DBFR 391

Search completed: February 29, 2004, 14:43:57

Job time : 57.0584 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:33:49 ; Search time 24.9147 Seconds
(without alignments)
3837.172 Million cell updates/sec

title: US-09-941-947A-26
infect score: 1526
sequence: 1 LTVCAKHVHLTGISAEQLL.....HSTYQLFIQAWFKGLAAVS 303

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

al number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1518	99.5	303	2 Q8GCS4	Q8GCS4 pantoea ste
2	1295.5	84.9	302	2 Q47842	Q47842 pantoea agg
3	1182.5	77.5	303	2 Q8VUJ8	Q8VUJ8 pantoea agg
4	490	32.1	359	2 Q93C19	Q93C19 xanthobacte
5	425	27.9	295	2 P94788	P94788 flavobacter
6	412.5	27.0	294	16 Q9K969	Q9K969 bacillus ha
7	399	26.1	316	2 Q9KIX4	Q9KIX4 bradyrhizob
8	391	25.6	299	16 Q8XX94	Q8XX94 raietonia s
9	385.5	25.3	367	10 Q9SX26	Q9SX26 daucus caro
10	377.5	24.7	347	10 Q9ZU77	Q9ZU77 arabidopsis
11	376.5	24.7	293	16 Q8NWD6	Q8NWD6 staphylococ
12	376.5	24.7	357	10 Q9L1A0	Q9L1A0 arabidopsis
13	375.5	24.6	293	16 Q9RTY2	Q9RTY2 staphylococ
14	369.5	24.2	369	10 Q8LSC4	Q8LSC4 cistus inca
15	368	24.1	360	10 Q9LHR4	Q9LHR4 arabidopsis
16	365.5	24.0	303	16 Q7W7Q1	Q7W7Q1 bordetella

17	364	23.9	376	10	Q04046	arabidopsis
18	363.5	23.8	303	16	Q7W138	bordetella
19	363.5	23.8	303	16	Q7VW86	bordetella
20	362.5	23.7	383	10	Q8W1R9	abies grand
21	361	23.8	372	10	Q8W1R9	abies grand
22	360.5	23.6	294	16	Q9AGM4	caulobacter
23	360	23.6	291	2	Q9RLH2	paracoccus
24	360	23.6	370	10	Q94ID7	hevea bras
25	359.5	23.6	343	10	Q22043	arabidopsis
26	359.5	23.6	356	10	Q81099	helianthus
27	359.5	23.6	360	10	Q9LUB1	arabidopsis
28	359	23.5	305	16	Q98FG7	rhizobium 1
29	359	23.5	371	10	Q8LAW5	arabidopsis
30	358.5	23.5	335	16	Q8UBX7	agrobacteri
31	358.5	23.5	336	10	Q9XIC0	arabidopsis
32	358.5	23.5	366	10	Q7X192	oryza sativ
33	357.5	23.4	304	16	Q8FYT3	brucella su
34	357.5	23.4	360	10	Q9LUD9	arabidopsis
35	357.5	23.4	393	10	Q9ZPM3	taxus canad
36	356.5	23.4	304	16	Q8YJ16	bruceella me
37	355.5	23.3	360	10	Q8LXN9	arabidopsis
38	355	23.3	378	10	Q38917	arabidopsis
39	354	23.2	302	16	P72683	synechocyst
40	353.5	23.2	360	10	Q8GY09	arabidopsis
41	352.5	23.1	368	10	Q9SSU0	croton subl
42	352	23.1	309	16	Q8Z085	anabaena sp
43	351.5	23.0	377	10	Q9SER3	mentha pipe
44	349	22.9	309	16	Q7U881	synechococ
45	347	22.7	293	16	Q8CSG3	staphylococ

ALIGNMENTS

RESULT 1

Q8GCS4	PRELIMINARY;	PRT;	303 AA.
AC	Q8GCS4		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Geranylgeranyl pyrophosphate synthase.		
GN	CETE.		
OS	Pantoea stewartii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Pantoea.		
OX	NCBI_TaxID=66269;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 8200;		
RA	deSouza M.L., Kollmann S.R., Schroeder W.A.;		
RT	"Carotenoid Biosynthesis (MO 02/079395 A2)";		
RL	Submitted (OCT-2002) to the EMBL/GenBank/DDBB databases.		
DR	EMBL; AY166713; AAN85596.1;		
DR	GO; GO:0008299; P:isoprenoid biosynthesis; IEA.		
DR	InterPro; IPR000092; Polyprenyl synth.		
DR	InterPro; IPR008949; Terpenoid synth.		
DR	PFAM; PF00348; polyprenyl synth_1		
DR	PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.		
DR	PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.		
SQ	SEQUENCE 303 AA; 32430 MW; 5CBEF868FD7B432B CRC64;		
Query Match 99.5%; Score 1518; DB 2; Length 303;			
Best Local Similarity 99.3%; Pred. No. 1.4e-114;			
Matches 301; Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
QY	1 LTVCAKHVHLTGISAEQLLADISRLDQLLPVQGERDCVGAAMEEGTLAPKRIKRLPMLL 60		
Db	1 MWCAKHVHLTGISAEQLLADISRLDQLLPVQGERDCVGAAMEEGTLAPKRIKRLPMLL 60		
QY	61 LITARDLGCAISHGGLLDLACAVEMVHAASLILDDMPQMDAQMERGRPTIHTQYGEHVA 120		
Db	61 LITARDLGCAISHGGLLDLACAVEMVHAASLILDDMPQMDAQMERGRPTIHTQYGEHVA 120		

1 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 2 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 3 Geranylgeranyl diphosphate synthase.
 4 CRTE.
 5 Xanthobacter sp. (strain Py2).
 6 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 7 Hyphomicrobiales; Xanthobacter.
 8 NCBI_TaxID=78245;
 9 [1]
 10 SEQUENCE FROM N.A.
 11 STRAIN=Py2;
 12 Larsen R.A.; Metcalf W.W.;
 13 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 14 EMBL; AF408847; AAL01998.1;
 15 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
 16 InterPro; IPR000092; Polyprenyl synth.
 17 InterPro; IPR008949; Terpenoid synth.
 18 Pfam; PF00348; Polyprenyl synth_1
 19 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 20 SEQUENCE 359 AA; 37348 MW; AE33F9E22D4B727D CRC64;
 21
 22 Query Match 32.1%; Score 490; DB 2; Length 359;
 23 Best Local Similarity 44.9%; Pred. No. 2.2e-31;
 24 Matches 120; Conservative 41; Mismatches 88; Indels 18; Gaps 7;
 25
 26 19 LIADISRLDQLLPVQGERDCV-GAAMREGLTAPGKRIIRPMLLLLTARDLGCALSHGGLL 77
 27 71 LRAVIDRRLGLLVPAAASHPAVLHAMRHILLSPGKRLRPLTLAAAIQENA--SENAVL 128
 28 78 DLACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGVIAE 137
 29 129 DFGCALEMHAASLIIDVDDLPMDDAAMRAQPTTHVQYGEDVAVLASIGLLSRAFGVAAA 188
 30 138 ARLGLTPIAKTRAVSELSTAGHGLVQGFQKDL--SEGDKPSADAILLTQPKTSTLFC 195
 31 189 AFGVSEARLEAVILSAVSGSLGCGQYDRLPSG---RSLSATEDVNRRTGVLPS 245
 32 195 ASTQMASIAANASCEARENHFRSLDLGOAFOLLDDLTGKMTDT---GKQINQDAGKSTL 252
 33 246 AVEIAGHVAADFTQGHKALAGHVGRAVQILDDIADASSASLSGKGVGDHAKATV 305
 34 253 VNLLSGAVVEERLQHLRLASEHLSAA 279
 35 306 IASLGA---PRAR---KLLSEHLGA 325
 36
 37 RESULT 5
 38 P94788 PRELIMINARY; PRT; 295 AA.
 39 P94788;
 40 01-MAY-1997 (TrEMBLrel. 03, Created)
 41 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 42 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 43 Geranylgeranyl synthase.
 44 CRTE.
 45 S Flavobacterium sp. ATCC 21588.
 46 Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;
 47 Flavobacteriaceae; Flavobacterium.
 48 NCBI_TaxID=50286;
 49 [1]
 50 SEQUENCE FROM N.A.
 51 STRAIN=R1534;
 52 MEDLINE=97186694; PubMed=9034310;
 53 Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
 54 van Loon A.P.;
 55 "Isolation and characterization of the carotenoid biosynthesis genes
 56 of Flavobacterium sp. strain R1534.";
 57 Gene 185:35-41(1997).
 58 EMBL; U62808; AAC44848.1;
 59 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
 60 InterPro; IPR000092; Polyprenyl synth.
 61 InterPro; IPR008949; Terpenoid synth.
 62 Pfam; PF00348; polyprenyl_synth_1.
 63 R

DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 SQ SEQUENCE 295 AA; 31332 MW; 849AC89EBCABC01 CRC64;
 1
 2 Query Match 27.9%; Score 425; DB 2; Length 295;
 3 Best Local Similarity 39.6%; Pred. No. 3e-26;
 4 Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
 5
 6 23 IDSRLDQLLPVQGERDCV-----GAAMREGLTAPGKRIIRPMLLLLTARDLGCALSHGGL 75
 7 13 VEIRLAQ---ISQFGVWSAPLGAAMSDAALSPGKRFRAVLMAVAE-----SSGVCVD 63
 8 76 -LLDLACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134
 9 64 AMVDAACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 123
 10 135 IAAEGLTPIAKTRAVSELSTAGHGLVQGFQKDLSEGDKPSADAILLTQPKTSTLFC 194
 11 124 LGEARGATPDQARLVASMSRAMGPVGLCAGQDLDL--HAPKDAAGIEREQDLTKTVLFL 180
 12 195 CASTQMASIAANASCEARENHFRSLDLGOAFOLLDDLTGKMTDT---GKQINQD-AGKS 250
 13 181 VAGLEMLSLIKGLDKRAETQLMAFGRLGRVQSVDDLLDVIGKASTGKTARDTAPG 240
 14 251 TLVNLIGSGAVBERLQHLRLASEHLSAACQNGHSTTQLF 290
 15 241 PKGGLMAVQMGD-VAQHYRASAZQDLMLR-----TRLF 274
 16
 17 RESULT 6
 18 Q9K969 PRELIMINARY; PRT; 294 AA.
 19 ID Q9K969;
 20 01-OCT-2000 (TrEMBLrel. 15, Created)
 21 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 22 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 23 Geranyltransferase.
 24 BH2781.
 25 OS Bacillus halodurans.
 26 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 27 NCBI_TaxID=86665;
 28 [1]
 29 SEQUENCE FROM N.A.
 30 STRAIN=C-125 / JCM 9153;
 31 MEDLINE=20512582; PubMed=11058132;
 32 Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 33 Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 34 Horikoshi K.;
 35 "Complete genome sequence of the alkaliphilic bacterium Bacillus
 36 halodurans and genomic sequence comparison with Bacillus subtilis.";
 37 Nucleic Acids Res. 28:4317-4331(2000).
 38 EMBL; AP001516; BAB06500.1;
 39 PIR; E83997; E83997.
 40 GO; GO:0016740; P:transferase activity; IEA.
 41 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
 42 GO; GO:000092; Polyprenyl synth.
 43 InterPro; IPR008949; Terpenoid synth.
 44 InterPro; IPR00348; polyprenyl synth_1.
 45 Pfam; PF00348; polyprenyl synth_1.
 46 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 47 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 48 Transferase; Complete proteome.
 49 SQ SEQUENCE 294 AA; 32653 MW; 3B2942868A62AD9E CRC64;
 50
 51 Query Match 27.0%; Score 412.5; DB 16; Length 294;
 52 Best Local Similarity 37.6%; Pred. No. 3.1e-25;
 53 Matches 103; Conservative 47; Mismatches 113; Indels 11; Gaps 4;
 54
 55 18 QLIADISRLDQLLPVQGER-----DCVGAAMREGLTAPGKRIIRPMLLLLTARDLGCALSH 73
 56 7 QFLDEIKDIIEERMPAHLIERLSNPMVLSKAGKRIIRPALLATWKSFKQDISQ 66
 57 74 GGLLDLACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFG 133
 58 67 G--IDLACAEIMTYSLIHDLPSMDDDDIRRGKPTNHKVFGEHAHLAGDALLTYSFE 124


```

134 VIAAEGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAAILTNQPKSTL 193
125 IVAKMGVDPAKTLLEELARAAGPEGMGVGVADIEGNGKLTVEGLEVIHHTKCAL 184
194 FCASQTWASTANANASCEARENLRHPSLDLGGAFOLLDLTDGKTD---TGKDINODAG-- 248
185 LSFAIVAGARLAASQDNIENIRFERELGELLPQIKDDILDVEGDAALGKPVGSDGNG 244
249 KSTLVNLVLSGGAVERLRQHRLASHLASAACON 282
245 KSTYPSLLTLEGAKELHLLTLLAKEVLYSVQVN 278

RESULT 7
9KIX4
ID Q9KIX4 PRELIMINARY; PRT; 316 AA.
AC Q9KIX4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Geranylgeranyl synthase.
EN CRTE.
OS Bradyrhizobium sp. ORS278.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
NCBI_TaxID=114615;
[1]
SEQUENCE FROM N.A.
STRAIN=ORS278;
MEDLINE=20309720; PubMed=10851005;
RA Hannibal L., Lorquin J., Angles d'Ortoli N., Garcia N.,
RA Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;
RA "Isolation and characterization of the canthaxanthin biosynthesis
RA genes from the photosynthetic bacterium Bradyrhizobium sp. strain
RA ORS278.";
RA J. Bacteriol. 182:3850-3853(2000).
AL EMBL; AF218415; AAF78199.1;
DR GO: 0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth. 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
SQ SEQUENCE 316 AA; 33713 MW; 9921ABD798A3FF16 CRC64;

Query Match 26.1%; Score 399; DB 2; Length 316;
Best Local Similarity 37.2%; Pred. No. 4.2e-24;
Matches 109; Conservative 38; Mismatches 106; Indels 40; Gaps 7;

2Y 7 KXVHLTGISBQLLADISRLDQLLPVQGRDCVGAAMREGT-----LAPG 52
DB 3 KPVDLTDTAA-----FETQLDRWRGRIGE--AVAEAMAFGTTVPAPLQAGMASHVLGG 54
2Y 53 KRIRPMLLLTARDLGCAISHGGLDLACAVEMVHAASLIDMPCMDDDAQMRRGRPTIH 112
DB 55 KRYRGMVLVAGSLG--VPEQLSSAVAETTHAASLVVDLPAMDARRRSQPAHE 112
2Y 113 TOYGEHVAILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDL-- 170
DB 113 VAFGEATILSSIALIAPAMEWVARDQLSPASRSSIVDTLSHAIGPQALCGGQYDLYP 172
2Y 171 ----SEGKPRSDAAILTNQPKSTLFCASQTWASTANANASCEARENLRHPSLDLGGAF 226
DB 173 PYATEQD-----LHRYQRKTSALFVAARFCTALLAEVDPETLLIRIARAGRLGVAF 225
2Y 227 QLLDPLTDGKTD---GRDINQDAGKSTLVNLVLSGGAVERLRQHRLASEHL 276
DB 226 QIFDDLLDLTGDAHAGKVDQDGHGTVTLATLLGPARAERAADELAAVQKEL 278

RESULT 8
9KXX94
ID Q8KXX94 PRELIMINARY; PRT; 299 AA.

```

```

AC Q8XX94;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable geranyltransferase (Farnesyl-diphosphate synthase)
DE protein (EC 2.5.1.10).
EN ISPA OR RSC2222 OR RS01377.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
NCBI_TaxID=305;
[1]
SEQUENCE FROM N.A.
STRAIN=GMI1000;
MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguanave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunhac S., Denange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin M., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Welssenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RA Nature 415:497-502(2002).
RL EMBL; AL646068; CAD15929.1;
DR GO: 0004337; F:geranyltransferase activity; IEA.
DR GO: 0016740; F:transferase activity; IEA.
DR GO: 0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth. 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 299 AA; 30911 MW; 7BF97371A1701382 CRC64;

Query Match 25.6%; Score 391; DB 16; Length 299;
Best Local Similarity 38.4%; Pred. No. 1.7e-23;
Matches 108; Conservative 43; Mismatches 104; Indels 26; Gaps 8;

QY 19 LEADIDSLDQLP---VOGRDCVGAAMREGTLAQKRIKRPMLLLLTARDLGCAISHGG 75
DB 11 VVARTESALERALPGESVAPQR--LHAAMEYATLGAQKRVPR--LLAHAAGALGEA--SPEA 66
QY 76 LLDLACAVEMVHAASLIDMPCMDDDAQMRRGRPTIHTOYGEHVAILAAVALLSKAFGI 135
DB 67 LDGVSCAVEMVHAASLIDMPCMDDDDLRGRPTVHRAVDEATALLVGDALQTOAFIAL 126
QY 136 AEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAAILTNQPKSTLFC 195
DB 127 AELGAVSPATRAGLVGELARASGSLGMAGGQALDLQSGVVALSQDALETMRHMKTALLR 186
QY 196 ASTOMASTANANASCEARENLRHPSLDLGGAFOLLDLTDGKTD---GKDINODAG 248
DB 197 ASLRMGALCAGVNAARLEQVDAYAGAVGLAFQVDDILDTATATLGTAGKDEAND-- 244
QY 249 KSTLVNLVLSGGAVERLRQHRLASEHLASAACQNGHSTTQL 289
DB 245 KPTVYSILG-----LERALADALHAAA--GAVAQL 275

RESULT 9
Q9SXZ6
ID Q9SXZ6 PRELIMINARY; PRT; 367 AA.
AC Q9SXZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GPP synthase.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Apiales; Apiaceae; Scandiceae; Daucinae;
OC Daucus.

```

```

1 NCBI_TaxID=4039;
2 [1] SEQUENCE FROM N.A.
3 TISSUE=Root;
4 "Daucus carota GDDP synthase 2.";
5 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
6 EMBL; AB027706; BAA78047.1;
7 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
8 InterPro; IPR000922; Polyrenyl synth.
9 Pfam; PF00348; polyrenyl synth; 1.
10 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
11 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
12 SEQUENCE 367 AA, 40045 MW, 44DB284991A05E3 CRC64;

Query Match 25.3%; Score 385.5; DB 10; Length 367;
Best Local Similarity 37.2%; Pred. No. 6.3e-23;
Matches 99; Conservative 49; Mismatches 99; Indels 19; Gaps 7;

23 IDSRLDQLPVGGRDCVGAAREGTLAPGKRIRPMLLLTARDLGAISHGGLDACA 78
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
87 INKALDEAVFQ-KPAITHEAMRYSLLAGGRVPLCIAS-----CELV-GLSWAMP 139
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
79 LACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAPGVIAEA 138
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
140 MACALEMTWMSLIHDDLPCMDNDLRRGKPTNHKVFGEVAVLAGDALISPAPEHLAT 199
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
139 -EGLTPIAKTRAVSELSTAIGMGLVQSGKDLSEGDGKPSADAILTNQKSTLFCAS 197
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
200 TMTVTPGRVFAIELSGAVSGVLVAGQIMDCNQGKVDVLDLEIYHVKHSKLEAS 259
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
198 TOMASIAANASCEARENLRHPSLDLQGFQALLDLDTGMDTGDINQDAG-----KST 251
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
260 VVCGAILGGGNVIEVRKTKYKARCGILGFQVVDIILD-VTKSSAEAGKTAGKOLVTDKTT 318
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
252 LVNKLGSVAVEELPQHLRLASEHLS 277
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
319 YPKMLGERAKRPATELVAQAEELS 344
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 10
12U77
2 Q9ZU77 PRELIMINARY; PRT; 347 AA.
3 01-MAY-1999 (TRENBLrel. 10, Created)
4 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
5 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
6 Putative geranylgeranyl pyrophosphate synthase.
7 AT2G18620
8 Arabidopsis thaliana (Mouse-ear cress).
9 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
10 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
11 eurosids II; Brassicales; Brassicaceae; Arabidopsis.
12 NCBI_TaxID=3702;
13 [1] SEQUENCE FROM N.A.
14 STRAIN=cv. Columbia;
15 MEDLINE=20083487; PubMed=10617197;
16 Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
17 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
18 Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
19 Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
20 Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Sonerville C.R.,
21 Copenhagen G.B., Preuss D., Nierman W.C., White O., Eisen J.A.,
22 Salzberg S.L., Fraser C.M., Venter J.C.;
23 "Sequence and analysis of chromosome 2 of the plant Arabidopsis
24 thaliana.";
25 Nature 402:761-768(1999).
26 [2]
27 SEQUENCE FROM N.A.
28 STRAIN=cv. Columbia;
29 Lin X.;
```

```

1 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
2 EMBL; AC006135; RAD12206.1;
3 FIR; B84566; B84566.
4 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
5 InterPro; IPR000922; Polyrenyl synth.
6 Pfam; PF00348; polyrenyl synth; 1.
7 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
8 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
9 SEQUENCE 347 AA, 37390 MW, CC68CD1D22F8840 CRC64;

Query Match 24.7%; Score 377.5; DB 10; Length 347;
Best Local Similarity 37.6%; Pred. No. 2.6e-22;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

23 IDSRLDQLPVGGRDCVGAAREGTLAPGKRIRPMLLLTARDLGAISHGGLDACA 82
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
70 LBSAVSLREPDK-----IHEAIRYSLLAGRVRPVLCAACELVGGESVA--LPAACA 122
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
83 VEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAPGVIAEASGLT 142
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
123 VEMIHTMSLIHDDLPCMDNDLRRGKPTNHKVFGEVAVLAGDALISPAPEHLATSTAVS 182
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
143 PIATKTRAVSELSTAIGMGLVQSGKDLSEGDGKPS-----ADAILLTNQKSTLFCAS 199
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
183 PARVYRAIGELAKAIGSKGLVAGQVVDLTSGMDNDVGLVLEFIHVHKTAVLELATV 242
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
200 MASIAANASCEARENLRHPSLDLQGFQALLDLDTGMDTGDINQDAGKSTLVNLL 256
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
243 LGAIVGGSGDEVEKLRERFARCIGILGFQVVDIILD-VTKSSEELGKTAGKOLIADKL 298
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 11
12U77
2 Q8NWD6 PRELIMINARY; PRT; 293 AA.
3 01-OCT-2002 (TRENBLrel. 22, Created)
4 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
5 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
6 IspA protein.
7 GN ISPA OR MW1474.
8 Staphylococcus aureus (strain MW2).
9 Bacteria; Firmicutes; Bacillales; Staphylococcus.
10 NCBI_TaxID=196620;
11 [1] SEQUENCE FROM N.A.
12 MEDLINE=22040717; PubMed=12044378;
13 Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
14 Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
15 Yamanoto K., Hiramatsu K.;
16 "Genome and virulence determinants of high virulence community-
17 acquired MRSA.";
18 Lancet 359:1819-1827(2002).
19 EMBL; AF004827; BAB95339.1;
20 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
21 InterPro; IPR000922; Polyrenyl synth.
22 InterPro; IPR008949; Terpenoid synth.
23 Pfam; PF00348; polyrenyl synth; 1.
24 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
25 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
26 Complete proteome.
27 SEQUENCE 293 AA; 32778 MW; 4A1920D061BB2CA7B CRC64;

Query Match 24.7%; Score 376.5; DB 16; Length 293;
Best Local Similarity 35.5%; Pred. No. 2.5e-22;
Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5;

11 LTGIGAFQALLADISRLDQLPVGGRDCVGAAREGTL-----APGKIRPMLLLTARD 66
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1 MTNLEPNKLIIDEVNNLS-----VAINKSVNDTQLESMLYSLNAGGKRIRPVLILLDS 56
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
67 LGCAJSHGGLLDLACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVA 126
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
Db 57 LNTVEYELG--MKSAIALEMIHTYSLIHDDLPAMDNDYRRGKLTNHNKVKYGEWTAIAGDA 114
127 LLSKAFGVIAEAGELTPIATRAVSELSTAIGVQGLVQGFQKDLSEGDKPRGADAILLTN 186
115 LITKAFELISDDRLTDEVKIKVLQRLSTIASGHVGVGGQMLDMQSEGQPIDLTLEMIH 174
187 QFTKSTLFCASTQWASIAANASCEARENLRHPSLDLQGAQFOLLDDLTGMDT---TGKDI 243
175 KTKTGALLTFVMSAADIANVDDATKEHLESYSYHLGMFMFQIKDDLLDCYGEAKLGKV 234
244 NQDA--GKSTLVNLGSGAVEERLRQHLRLASEHLS 277
235 GSDLENNKSTVSLGKGAGDKLTYHRDAAVDEL 270

RESULT 12
29LIAO PRELIMINARY; PRT; 357 AA.
ID Q9LIAO
AC Q9LIAO;
YT 01-OCT-2000 (TReMBLrel. 15, Created)
YT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
YT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
JE Geranylgeranyl pyrophosphate synthase.
S Arabidopsis thaliana (Mouse-ear cress).
X Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
X Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids;
X eurosids II; Brassicales; Brassicaceae; Arabidopsis.
XX NCBI_TaxID=3702;
XN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RC Kameko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RZ Nakamura Y.;
RA "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP001309; BAB02589.1; -.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth_1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
DR TRANSFERASE; Complete proteome.
SQ SEQUENCE 357 AA; 38502 MW; 41A40566CC8962F3 CRC64;

Query Match 24.7%; Score 376.5; DB 10; Length 357;
Best Local Similarity 37.1%; Pred. No. 3.3e-22;
Matches 103; Conservative 50; Mismatches 84; Indels 41; Gaps 10;

QY 23 IDSRLDQLLPVQGERDCVGAAREGLTAPGRIRPMLL-----LTARDLGCALSHGGLLD 78
DB 74 VSAALNVSVLQ-EPLTIQEAIVYSLLAGGRVRLPLLCIAACELVGGDEATAMS----- 126
QY 79 LACAVEMVHARSLIIDDMPGMDADQVRGRPTHTQYGEHVAILAAVALLSKAF----- 132
DB 127 AACAVEMHITSLIHDDLPAMDNDYRRGKLTNHNKVKYGEWTAIAGDA 114
QY 133 --GVIAEAGELTPIATRAVSELSTAIGVQGLVQGFQKDL-SEGDKPRSA--DAILLTNQ 187
DB 127 AACAVEMHITSLIHDDLPAMDNDYRRGKLTNHNKVKYGEWTAIAGDA 114
QY 187 QFTKSTLFCASTQWASIAANASCEARENLRHPSLDLQGAQFOLLDDLTGMDT---TGKDI 243
DB 187 KTKTGALLTFVMSAADIANVDDATKEHLESYSYHLGMFMFQIKDDLLDCYGEAKLGKV 234
QY 248 GKST-----LVNLLGSGAVEERLRQHLRLASEHL 276
DB 241 HKTAALEAAAVGAMGGGTETEEKURKARCTGLLPQVDDILD-VTSTKELGKTA 299
QY 248 GKST-----LVNLLGSGAVEERLRQHLRLASEHL 276
DB 241 HKTAALEAAAVGAMGGGTETEEKURKARCTGLLPQVDDILD-VTSTKELGKTA 299
```

```
Db 300 GKQVMAGKLTYPRLIGLERSREVAEKLR---REARQL 334

RESULT 13
Q99TX2 PRELIMINARY; PRT; 293 AA.
ID Q99TX2
AC Q99TX2;
YT 01-JUN-2001 (TReMBLrel. 17, Created)
YT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
YT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE ISPA protein (Geranyltransferase homolog).
GN ISPA OR SAV1521 OR SAV1352.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
XN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Ogasawara N., Hayashi H., Hiramatsu K.;
RA "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57683.1; -.
DR EMBL; AP003362; BAB57683.1; -.
DR PIR; A89932; A89932.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth_1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
DR TRANSFERASE; Complete proteome.
SQ SEQUENCE 293 AA; 32746 MW; E7AFEBBBA868F7D8 CRC64;

Query Match 24.6%; Score 375.5; DB 16; Length 293;
Best Local Similarity 35.5%; Pred. No. 3e-22;
Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5;

QY 11 LTGISAEOLLADISRLDQLLPVQGERDCVGAAREGLT-----APGRIRPMLLLTARD 66
DB 11 MNLPMKLIIDEVNNLS-----VAINKSVMDTQLEESMLYSLNAGGKIRPVLILLTDS 56
QY 67 LGCALSHGGLLDLACAVEMVHARSLIIDDMPGMDADQVRGRPTHTQYGEHVAILAAVA 126
DB 57 LNTVEYELG--MKSAIALEMIHTYSLIHDDLPAMDNDYRRGKLTNHNKVKYGEWTAIAGDA 114
QY 127 LLSKAFGVIAEAGELTPIATRAVSELSTAIGVQGLVQGFQKDLSEGDKPRGADAILLTN 186
DB 115 LITKAFELISDDRLTDEVKIKVLQRLSTIASGHVGVGGQMLDMQSEGQPIDLTLEMIH 174
QY 187 QFTKSTLFCASTQWASIAANASCEARENLRHPSLDLQGAQFOLLDDLTGMDT---TGKDI 243
DB 175 KTKTGALLTFVMSAADIANVDDATKEHLESYSYHLGMFMFQIKDDLLDCYGEAKLGKV 234
QY 244 NQDA--GKSTLVNLGSGAVEERLRQHLRLASEHLS 277
DB 235 GSDLENNKSTVSLGKGAGDKLTYHRDAAVDEL 270

RESULT 14
Q9LSC4 PRELIMINARY; PRT; 369 AA.
ID Q9LSC4
AC Q9LSC4;
YT 01-OCT-2002 (TReMBLrel. 22, Created)
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 8.43386 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947A-26
Project score: 1526
Sequence: 1 LTVCAKXVHLTGISAQELL.....HSTTQLFIQAWFDKLAWS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 78:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1344.5	88.1	302	2 A37802	crTE protein - Erw
2	1295.5	84.9	302	2 S52584	crTE protein - Erw
3	752	49.3	308	2 C39273	phytoene synthase
4	412.5	27.0	234	2 E83997	geranyltranstransf
5	377.5	24.7	347	2 E84566	probable geranylge
6	375.5	24.6	293	2 A89932	hypothetical prote
7	372	24.4	369	2 S53722	farnesyltranstrans
8	369	24.2	371	2 P85434	geranylgeranyl pyr
9	364.5	23.9	366	2 T10452	farnesyltranstrans
10	364	23.9	376	2 T02429	geranylgeranyl pyr
11	361	23.7	372	2 G84566	probable geranylge
12	360.5	23.6	294	2 D87505	geranyltranstransf
13	360.5	23.5	297	2 JX0257	geranyltranstransf
14	358.5	23.5	304	2 F97685	geranyltranstransf
15	358.5	23.5	335	2 AH2910	geranyltranstransf
16	356.5	23.4	304	2 A13285	geranyltranstransf
17	355	23.3	378	2 S71230	geranylgeranyl pyr
18	354	23.2	302	2 S74538	geranylgeranyl pyr
19	352	23.1	309	2 AE1833	geranylgeranyl dip
20	346.5	22.7	272	2 A61961	geranyltranstransf
21	345.5	22.6	316	2 T11021	farnesyltranstrans
22	345.5	22.6	357	2 T09966	farnesyltranstrans
23	339	22.2	294	2 A82267	geranyltranstransf
24	335.5	22.0	306	2 AH0385	geranyltranstransf
25	332	21.8	299	2 JQ0665	geranyltranstransf
26	329	21.6	293	2 AG1607	geranyltranstransf
27	328.5	21.5	298	2 E81796	geranyltranstransf
28	328	21.5	299	2 C90688	geranyltranstransf
29	328	21.5	299	2 G85538	geranyltranstransf

geranyltranstransf
y4KU protein - Rhi
probable dimethyla
geranyltranstransf
farnesyltranstransf
geranyltranstransf
farnesyltranstransf
geranyltranstransf
geranylgeranyl pyr
geranyltranstransf
farnesyltranstransf
geranyltranstransf
geranyltranstransf
phytoene synthase
geranyltranstransf
geranyltranstransf

30 322 21.1 299 2 AG0554
31 320 21.0 332 2 T10875
32 316 20.7 332 2 T40213
33 314 20.6 293 2 AC1245
34 313.5 20.5 300 2 T06969
35 313 20.5 291 2 D82778
36 313 20.5 326 2 T05674
37 309.5 20.3 295 2 C64123
38 308 20.2 288 2 S49625
39 307 20.1 259 2 F81217
40 307 20.1 288 2 T50750
41 305.5 20.0 291 2 F98007
42 303.5 19.9 291 2 G95139
43 301.5 19.8 295 2 F81139
44 297.5 19.5 289 2 S04407
45 296 19.4 281 2 B81261

ALIGNMENTS

RESULT 1

A37802
crTE protein - Erwinia uredovora
C:Species: Erwinia uredovora
C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
C:Accession: A37802
R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashir
J. Bacteriol. 172, 6704-6712, 1990
A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by function
A:Reference number: A37802; MOID:9107214; PMID:2254247
A:Accession: A37802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <MIS>
A:Cross-references: GB:D90087; NID:G216681; PIDN:BAAL4124.1; PID:G216682
C:Superfamily: Geranyltransferase

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 1.7e-104;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

QY 1 LTVCAKXVHLTGISAQELLADIDSLDQLLPVQGERDCVGAAMREGTTLAPGKRIRPMLL 60
DB 1 MTVCAKXVHLTFDAEQALLADIDRLDQLLPVGERDVVGAAMREGALAPGKRIRPMLL 60
QY 61 LITARDLGCAISHGGLLDLACAVEVWHAASLILDDMPQDDQMRGRPTIITQGEHVA 120
DB 61 LITARDLGCAVSHDGLDLACAVEVWHAASLILDDMPQDDQMRGRPTIITQGEHVA 120
QY 121 ILAVALLSKAFGVIAEAGLTETPIAKTRAVSELSAIGMGLVQGOFKDLSGDXPRSD 180
DB 121 ILAVALLSKAFGVIAADAGLTETPLAKNRVSELSAIGMGLVQGOFKDLSGDXPRSAE 180
QY 181 AILLTNQKSTLFCASQWASIAANASCEARENLFHSLDLGQAFQLDLDTGMDTGTG 240
DB 181 AILLTNHFKSTLFCASQWASIAANASCEARENLFHSLDLGQAFQLDLDTGMDTGTG 240
QY 241 KQINQDAGKSTLVNLLGSGAVERLRQHLRLASEHLSAACQGHSTTOLFQAWFDKLA 300
DB 241 KQSNQDAGKSTLVNLLGSPRAVERLRQHLRLQASEHLSAACQGHGA-TORPTQAWFDKLA 299
QY 301 AVS 303
DB 300 AVS 302

RESULT 2

S52584
crTE protein - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 24-Nov-1999
C:Accession: S52584

Rubin, Y.P.; Lai, B.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.
 Mol. Genet. 245, 417-423, 1994
 A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.
 A:Reference number: S52583; MUID:95107237; PMID:7808390
 A:Accession: S52584
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-302 <LIN>
 A:Cross-references: EXBL:M90698; NID:g148393; PIDN:AAA21260.1; PID:g148394
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
 C:Superfamily: geranyltransferase

Query Match 84.9%; Score 1295.5; DB 2; Length 302;
 Best Local Similarity 86.1%; Pred. No. 2e-100;
 Matches 261; Conservative 14; Mismatches 27; Indels 1; Gaps 1;
 QY 1 LTVCAKXVHLTGSABQLADISRLDQLPVQGERDCVGAAREGTLAPGKRIKRPML 60
 DB 1 MIVCAKXVPTSAQAQLADIDRLQQLPVQGERDPVGAAREGTLAPGKRIKRPML 60
 QY 61 LITARDLGCAISHGGLDLACAVEMVHAASLILDDMPCHDDAQMRGRPTIHTQYGEHVA 120
 DB 61 LITARDLGCAVSHGGLDLACAVEMVHAASLILDDMPCHDDAQIVGGRPTVHCQYGEHVA 120
 QY 121 ILAVALLSKARGVIRAEGLTPIAKTRAVSELSTAIGMCHVQCGFYDLSEGDKPRASD 180
 DB 121 ILAVALLSKARGVIRAEGLTPIAKTRAVSELSTAIGMCHVQCGFYDLSEGDKPRASAE 180
 QY 181 AILLTNOFTSTLFCASOMASIVANASERRDYLRPFLDLGQAFQLDLDTGMDTG 240
 DB 181 ALLTNEFTSTLFCASOMASIVANASERRDYLRPFLDLGQAFQLDLDTGMDTG 240
 QY 241 KQINDQKSTLVNLLGSAVERLRQHLRLASEHLSAACQNGHSTTQFIQAFDCKLA 300
 DB 241 KQINDQKSTLVNLLGSAVERLRQHLRLASEHLSAACQNGHSTTQFIQAFDCKLA 299
 QY 301 AVS 303
 DB 300 AVS 302

RESULT 3

C39273
 phytoene synthase - Erwinia herbicola
 C:Species: Erwinia herbicola
 C:Date: 18-Oct-1991 #sequence_revision 03-Feb-1994 #text_change 15-Oct-1999
 C:Accession: C39273; C33120
 R:Armstrong, G.A.; Alberti, M.; Hearst, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990
 A:Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in non
 A:Reference number: A39273; MUID:91088634; PMID:2263648
 A:Accession: C39273
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-308 <ARM>
 A:Cross-references: GB:M38424
 A:Note: the authors translated the codon CTC for residue 187 as Ala, ATT for residue 191
 e 211 as Ala
 A:Note: the authors inserted as Arg after residue 261 in figure 3
 R:Armstrong, G.A.; Alberti, M.; Hearst, J.E.
 submitted to the Protein Sequence Database, September 1990
 A:Reference number: A33120
 A:Accession: C33120
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-261,263-308 <AR2>
 A:Experimental source: strain Eho10, ATCC 39368
 C:Genetics:
 A:Gene: crtE
 C:Superfamily: geranyltransferase
 C:Keywords: carotenoid biosynthesis

Query Match 49.3%; Score 752; DB 2; Length 308;

Best Local Similarity 55.8%; Pred. No. 4e-55;
 Matches 159; Conservative 40; Mismatches 84; Indels 2; Gaps 2;
 QY 17 EOLLADIISRLDQLPVQGERDCVGAAREGTLAPGKRIKRPMLLTARDLGCAISHGGL 76
 DB 16 EVMROSIDHLAGLIPETDSQDIIVSLAVEGVMAPGKRIKRPMLLAARDLRQGSNPTL 75
 QY 77 LILACAVEMVHAASLILDDMPCHDDAQMRGRPTIHTQYGEHVAAILAVALLSKARGVIA 136
 DB 76 LILACAVELTTPASLMDMPCHDDAQMRGRPTIHTQYGEHVAAILAVALLSKARGVIA 135
 QY 137 BAEGTLPIAKTRAVSELSTAIGMCHVQCGFYDLSEGDKPRASDAILLTNOFTSTLFC 196
 DB 136 ATGDLPGERRAGVAVNELSTAVGQELVLGQPRDLNDAALDRTPDAILSTNHLKTLFSA 195
 QY 197 STOMASIAANASCEARENLRPFLDLGQAFQLDLDTGMDTGKINDQAGKSTLVNLL 256
 DB 196 MIVCAKXVPTSAQAQLADIDRLQQLPVQGERDPVGAAREGTLAPGKRIKRPML 255
 QY 257 GSGAV-EERLRQHLRLASEHLSAACQNGHSTTQFIQAFDCKLA 300
 DB 256 GADAARQKLRHIDSADKHLTFACPGQAIRO-FMELWFGHLLA 299
 RESULT 4
 E83997
 geranyltransferase BH2781 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
 C:Accession: E83997
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hir
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: E83997
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-294 <STO>
 A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA006500.1; GSPDB:GNO
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2781
 C:Superfamily: geranyltransferase

Query Match 27.0%; Score 412.5; DB 2; Length 294;
 Best Local Similarity 37.6%; Pred. No. 7.5e-27;
 Matches 103; Conservative 47; Mismatches 113; Indels 11; Gaps 4;
 QY 18 QLLADIDSLDQLPVQGER----DCVGAAREGTLAPGKRIKRPMLLTARDLGCAISH 73
 DB 7 QFLDIEIKDIIERRPAHIERLNSPDMKNSMLYSLKAGGKRIKRPMLLTATMKSFKQDISO 66
 QY 74 GGLDLACAVEMVHAASLILDDMPCHDDAQMRGRPTIHTQYGEHVAAILAVALLSKARG 133
 DB 67 G-IDLACAEIHTYSLIHDDLPSMDDDIRGKPTNHKVFCEAHAILAGDALLTYSFE 124
 QY 134 VIAEAGLTPIAKTRAVSELSTAIGMCHVQCGFYDLSEGDKPRASDAILLTNOFTSTL 193
 DB 125 IVAKKGVDPKATLCLIEELARAAGPGVGGVADIEGENOKLTVEGLEIYTHHRTKAL 184
 QY 194 FCASOMASIAANASCEARENLRPFLDLGQAFQLDLDTGMDTG---TKQINDQAG-- 248
 DB 185 LSPFAVAGARLADASEQDIENIRFRSRELGLLFOIKDDILDVEDGQAAIGKVPVSDGQNG 244
 QY 249 KSTLVNLLGSAVERLRQHLRLASEHLSAACON 282
 DB 245 KSTYPSLLTLEGAKEKHLHTLLAKKYLVSQVQN 278

RESULT 5

B84566
 probable geranyl pyrophosphate synthase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

```
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
;Accession: E84566
;Link: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
;Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
;us, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
;ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Status: preliminary
;Molecule type: DNA
;Residues: 1-347 <STO>
;Cross-references: GB:AE002093; NID:94217998; PIDN:AAD12206.1; GSPDB:GN00139
;Genetics:
;Gene: At2g18620
;Map position: 2
;Superfamily: geranyltransferase

Query Match 24.7%; Score 377.5; DB 2; Length 347;
Best Local Similarity 37.6%; Pred. No. 7.6e-24;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

Y 23 IDSRLDQLPVQGERDCVGAAMEEGTLAPKRIKIRPMLLLLTARDLGCAISHGGLDLACA 82
b 70 LGSVSLRBEPIK-----IHEAIRYSELARGKVRPVLCTAACELVGGEBSVA--LPAACA 122

Y 83 VEMVHAASLILDDMPQNDQACMERGPTTIHQYGEHVAILAAVALLSKAFGVIAEAGLT 142
b 123 VEMHTMSLTHDDLPCNDNDLARGKPTNHKIVGEDVAVLAGDLSFAFEHLATSTAVS 182

Y 143 PIAKTRAVSELSTAIGMOGLVQGVQFQKLSSEGDKPRS---ADAILLTNQPKTSTLFCASTQ 199
b 183 PARVWRAIGELAKIGSKGLVAGOVVDLISGGWDQNDVGLVLEFIHVHKTAVLEAATV 242

Y 200 MASTANASCEARENLRHFSLDLQGAFLDLDLTDGMDTKDINQDACKSTLVNLL 256
b 243 LGATVGGSGDEEVEKRRFARCIGLLFQVDDILD-VTKSSEELGTAGKDLIADKL 298

;RESULT 6
;Accession: A8932
;Species: Staphylococcus aureus
;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
;Accession: A8932
;Link: X.; Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino,
;na, A.; Mizutani-Ui, Y.; Kobayashi, K.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
;C.; Shiba, T.; Rattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
;ature 357, 1225-1240, 2001
;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
;Reference number: A89758; MUID:21311952; PMID:11418146
;Status: preliminary
;Molecule type: DNA
;Residues: 1-293 <KUR>
;Cross-references: GB:BA000019; PID:BA042614.1; GSPDB:GN00149
;Experimental source: strain N315
;Genetics:
;Gene: ispA

Query Match 24.6%; Score 375.5; DB 2; Length 293;
Best Local Similarity 35.5%; Pred. No. 9e-24;
Matches 98; Conservative 46; Mismatches 119; Indels 15; Gaps 5;

Y 11 LTGTSABQLADIDSLDQLLPVQGERDCVGAAMEEGLT-----APGKRIKIRPMLLLLTARD 66
b 1 MTNLPMNKLIDEVNNELS---VAINKSVMDTQLEESMLYSLNAGGKRIKIRPMLLLTLDLS 56

Y 67 LGCAISHGGLDLACAVEMVHAASLILDDMPQNDQACMERGPTTIHQYGEHVAILAAVA 126
b 57 LNTYEYELG--VKSAIALEMHTHTSLHDDLPEANDNDYRRGKLTNKHVGYEFTAILAGDA 114

Y 127 LLSNAFGVIAEAGLTPIAKTRAVSELSTAIGMOGLVQGVQFQKLSSEGDKPRGADAILTN 186
```

```
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
;Accession: E84566
;Link: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
;Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
;us, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
;ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Status: preliminary
;Molecule type: DNA
;Residues: 1-347 <STO>
;Cross-references: GB:AE002093; NID:94217998; PIDN:AAD12206.1; GSPDB:GN00139
;Genetics:
;Gene: At2g18620
;Map position: 2
;Superfamily: geranyltransferase

Query Match 24.7%; Score 377.5; DB 2; Length 347;
Best Local Similarity 37.6%; Pred. No. 7.6e-24;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

Y 23 IDSRLDQLPVQGERDCVGAAMEEGTLAPKRIKIRPMLLLLTARDLGCAISHGGLDLACA 82
b 70 LGSVSLRBEPIK-----IHEAIRYSELARGKVRPVLCTAACELVGGEBSVA--LPAACA 122

Y 83 VEMVHAASLILDDMPQNDQACMERGPTTIHQYGEHVAILAAVALLSKAFGVIAEAGLT 142
b 123 VEMHTMSLTHDDLPCNDNDLARGKPTNHKIVGEDVAVLAGDLSFAFEHLATSTAVS 182

Y 143 PIAKTRAVSELSTAIGMOGLVQGVQFQKLSSEGDKPRS---ADAILLTNQPKTSTLFCASTQ 199
b 183 PARVWRAIGELAKIGSKGLVAGOVVDLISGGWDQNDVGLVLEFIHVHKTAVLEAATV 242

Y 200 MASTANASCEARENLRHFSLDLQGAFLDLDLTDGMDTKDINQDACKSTLVNLL 256
b 243 LGATVGGSGDEEVEKRRFARCIGLLFQVDDILD-VTKSSEELGTAGKDLIADKL 298

;RESULT 6
;Accession: A8932
;Species: Staphylococcus aureus
;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
;Accession: A8932
;Link: X.; Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino,
;na, A.; Mizutani-Ui, Y.; Kobayashi, K.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
;C.; Shiba, T.; Rattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
;ature 357, 1225-1240, 2001
;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
;Reference number: A89758; MUID:21311952; PMID:11418146
;Status: preliminary
;Molecule type: DNA
;Residues: 1-293 <KUR>
;Cross-references: GB:BA000019; PID:BA042614.1; GSPDB:GN00149
;Experimental source: strain N315
;Genetics:
;Gene: ispA

Query Match 24.6%; Score 375.5; DB 2; Length 293;
Best Local Similarity 35.5%; Pred. No. 9e-24;
Matches 98; Conservative 46; Mismatches 119; Indels 15; Gaps 5;

Y 11 LTGTSABQLADIDSLDQLLPVQGERDCVGAAMEEGLT-----APGKRIKIRPMLLLLTARD 66
b 1 MTNLPMNKLIDEVNNELS---VAINKSVMDTQLEESMLYSLNAGGKRIKIRPMLLLTLDLS 56

Y 67 LGCAISHGGLDLACAVEMVHAASLILDDMPQNDQACMERGPTTIHQYGEHVAILAAVA 126
b 57 LNTYEYELG--VKSAIALEMHTHTSLHDDLPEANDNDYRRGKLTNKHVGYEFTAILAGDA 114

Y 127 LLSNAFGVIAEAGLTPIAKTRAVSELSTAIGMOGLVQGVQFQKLSSEGDKPRGADAILTN 186
```

```
;Date: 15-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
;Accession: S53722
;Link: X.; Badillo, A.; Steppuhn, J.; Derviere, J.; Camara, B.; Kuntz, M.
;Plant Mol. Biol. 27, 425-428, 1995
;Title: Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsa
;Reference number: S53722; MUID:95195169; PMID:788631
;Status: nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-369 <BAD>
;Cross-references: EMBL:X80267; NID:9643093; PIDN:CAA56554.1; PID:9643094
;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
;Genetics:
;Gene: GGPPS
;Superfamily: geranyltransferase
;Keywords: carotenoid biosynthesis; Chloroplast; chromoplast; transferase

Query Match 24.4%; Score 372; DB 2; Length 369;
Best Local Similarity 34.5%; Pred. No. 2.4e-23;
Matches 99; Conservative 52; Mismatches 108; Indels 28; Gaps 7;

Y 23 IDSRLDQLPVQGERDCVGAAMEEGTLAPKRIKIRPMLLLLTARDLGCAISHGGLDLACA 82
b 89 VNKALDEAIIIVK-EPHVIHEAMRYSLLAGGKVRPMLCTAACELVGG--NQENAMAAACA 145

Y 83 VEMVHAASLILDDMPQNDQACMERGPTTIHQYGEHVAILAAVALLSKAF-GVIAEAGLT 141
b 146 VEMHTMSLTHDDLPCNDNDLARGKPTNHKIVGEDVAVLAGDLSFAFEHLVNSTAGV 205

Y 142 TPIAKTRAVSELSTAIGMOGLVQGVQFQKLS-EGDKPRSADAILTNQPKTSTLFCASTQM 200
b 206 TPSRIVGVAELAKSICTEGLVAGVADIKCTGNASVSLTLEFIHVHKTAALESSVWL 265

Y 201 ASIANNASCEARENLRHFSLDLQGAFLDLDLTDGMDTKDINQDACKSTLVN----- 254
b 266 GAILGGGTNVEVEKRRFARCIGLLFQVDDILD-VTKSSEELGTAGKDLVVDKTTYPK 324

Y 255 LLGSGAVVEERLROHLRLASEHLSAAACQNGHSTTQLFQIAWFDKXKLA 301
b 325 LLGLEKAKEPAEALNREAKQQLG-----FDSKAA 355

;RESULT 8
;Accession: F85434
;Species: Arabidopsis thaliana [mouse-ear cress]
;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 24-Aug-2001
;Accession: F85434
;Link: X.; Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
;Nature 402, 769-777, 1999
;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
;Reference number: A85001; MUID:20083488; PMID:10617198
;Status: preliminary
;Molecule type: DNA
```



```

;Residues: 1-371 <STO>
;Cross-references: GB:NC_001268; NID:g7270630; PIDN:CHB80347.1; GSPDB:GN00140
;Genetics:
;Gene: AY936810
;Map position: 4
;Superfamily: Geranyltransferase

Query Match      24.2%; Score 369; DB 2; Length 371;
Best Local Similarity 38.8%; Pred. No. 4.2e-23;
Matches 93; Conservative 46; Mismatches 91; Indels 10; Gaps 6;

Y 23 IDSRLDQLLPVQGERDCVGAARREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA 82
b 87 VNKALDSAVPLR-EPLKHEAMRYSLLAGGKVRPVLCAACELVGGESBTA--MPAACA 143
Y 83 VEMVHAASLILDDMPQMDAQMGRGRTHTQYGEHVAIAAVALLSKAFGVIAEA---E 139
b 144 VEMHTMSLIHDDLPCMDNDLLRRKPTNEKVFEDVAVLAGDALLSFAPEHLASATSSD 203
Y 140 GLTPPTAKTRAVSELSTAIGMGLVQGFQKDL-SEGDKPRA--DAILLTNQFTSTLPCA 196
b 204 VVSPVRVVRVAVGELAKALGTBGLVAGQVVDISSEGLDNDVGLHLEFTHLHKTAALEA 263
Y 197 STQMASIAANASCEARENHLPFSLDLGOAFOLLDDLTGDMTDTGKINODAGKSTLWLL 256
b 264 SAVLGAIVGGSDDEIERLRFKFAICIGLLFQVVDIIL-VTKSKELGKTAGKDLADKL 322

RESULT 9
Geranyltransferase (EC 2.5.1.29) precursor, chloroplast - white mustard
;Alternate names: geranylgeranyl-diphosphate synthase; geranylgeranyl-pyrophosphate syn
;Species: Sinapis alba (white mustard)
;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
;Accession: T10452
;Author: Bonk, M.; Hoffmann, B.; von Lintig, J.; Schledz, M.; Al-Babili, S.; Hobeika, E.; Klein
;Title: Chloroplast import of four carotenoid biosynthetic enzymes in vitro reveals dif
;Reference number: Z17023; MUID:97433278; PMID:9288918
;Accession: T10452
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-366 <BON>
;Cross-references: EMBL:X98795; NID:g1419757; PIDN:CAA67330.1; PID:g1419758
;Genetics:
;Gene: GGPS
;Superfamily: geranyltransferase
;Keywords: carotenoid biosynthesis; chloroplast; transferase

Query Match      23.9%; Score 364.5; DB 2; Length 366;
Best Local Similarity 39.5%; Pred. No. 9.9e-23;
Matches 96; Conservative 42; Mismatches 88; Indels 17; Gaps 8;

Y 23 IDSRLDQLLPVQGERDCVGAARREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA-- 80
b 83 VNKALDSAVPLR-EPLKHEAMRYSLLAGGKVRPVLCAACELV--GGESLAMP 135
Y 81 --CAVEMVHAASLILDDMPQMDAQMGRGRTHTQYGEHVAIAAVALLSKAFGVIAEA 138
b 136 ARCVAEMHTMSLIHDDLPCMDNDLLRRKPTNEKVFEDVAVLAGDALLSFAPEHLASA 195
Y 139 EG--LTPTAKTRAVSELSTAIGMGLVQGFQKDL-SEGDKPRA--DAILLTNQFTSTL 193
b 196 TSSEVSPARVVRVAVGELAKALGTBGLVAGQVVDISSEGLDNNVGLHLEKPTHLHKTAA 255
Y 194 FCASQMASIAANASCEARENHLPFSLDLGOAFOLLDDLTGDMTDTGKINODAGKSTLV 253
b 256 LEASAVTGGTGGSDDEIERLRFKFAICIGLLFQVVDIIL-VTKSSQELGKTAGDLIA 314
Y 254 NLL 256
b 315 DKL 317

```

RESULT 10

```

T02429
Geranylgeranyl pyrophosphate synthase (EC 2.5.1.1) - Arabidopsis thaliana
;Alternate names: pregeranylgeranyl pyrophosphate synthase; protein F27L4.2; protein P;
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Feb-2001
;Accession: T02429; T02430; S84629
;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, May 1998
;Description: Arabidopsis thaliana chromosome II BAC F27L4 genomic sequence.
;Reference number: Z14658
;Accession: T02429
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-376 <ROU>
;Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17083.1; PID:g3152604
;Experimental source: cultivar Columbia
;Accession: T02430
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-336 <ROW>
;Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17084.1; PID:g3152605
;Experimental source: cultivar Columbia
;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Accession: B84629
;Status: preliminary
;Molecule type: DNA
;Residues: 1-376 <STO>
;Cross-references: GB:AE002093; NID:g3152604; PIDN:AAC17083.1; GSPDB:GN00139
;Genetics:
;Gene: At2g23800; F27L4.2; F27L4.3
;Map position: 2
;Introns: 184/3
;Superfamily: Geranyltransferase
;Keywords: transferase

Query Match      23.9%; Score 364; DB 2; Length 376;
Best Local Similarity 38.1%; Pred. No. 1.1e-22;
Matches 102; Conservative 45; Mismatches 97; Indels 24; Gaps 9;

Y 23 IDSRLDQLLPVQGERDCVGAARREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA 82
b 96 VNKALEENIPV-GEPLKHEAMRYSLLAGGKVRPVLCAACELVGG--QENAAAPACA 152
Y 83 VEMVHAASLILDDMPQMDAQMGRGRTHTQYGEHVAIAAVALLSKAFGVIAEGLT 142
b 153 VEMHTMSLIKDDLPCMDNDLLRRKPTTHKYVGGVAILSGGALLSLAFEHMTTAE--- 209
Y 143 PIAKTR---AVSELSTAIGMGLVQGFQKDL-SEGDKPRA--DAILLTNQFTSTLPCA 196
b 210 -ISSERWVAVRELARSIGTRGLVAGQAMDISEGLDNEVGLHLEPHIVHKTAVLETT 268
Y 197 STQMASIAANASCEARENHLPFSLDLGOAFOLLDDLT-----GMTDTGKINODAG 248
b 269 AAVLGAIVGGSDDEIERLRFKFAICIGLLFQVVDIIL-VTKSKELGKTAGKDL 325
Y 249 KSTLVNLLGSGAVBELRHOHLASEHL 276
b 326 KLTYPKLGLEKSKFVKELTKDAFQHL 353

RESULT 11
G84566
probable geranylgeranyl pyrophosphate synthase [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002

```



```

;Accession: G84566
;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.R.; Umayam, L.; Tallon, L.;
; Iss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C.
;ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Accession: G84566
;Status: preliminary
;Molecule type: DNA
;Residues: 1-372 <STO>
;Cross-references: GB:AE002693; NID:g4185130; PIDN:AA08933.1; GSPDB:GN00139
;Genetics:
;Gene: At2g18640
;Map position: 2
;Superfamily: geranyltransferase

Query Match      23.7%; Score 361; DB 2; Length 372;
Best Local Similarity 37.4%; Pred. No. 2e-22;
Matches 99; Conservative 47; Mismatches 101; Indels 18; Gaps 8;

/ 23 IISRLDOLLPWGGRDCVCANREGTLAPGKIRPMWLLLTARDLGCASHGG--LIDL 79
c 92 INRALDEA:PLI-EPLNIHKHAYAILAGGKVRPILC-----AACLVGGERLAIQA 145
/ 80 ACAVEMVHAASLIIDMPCWDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFGVIAAE 139
b 146 ACAVEMIRHTMSLIIKDLLPCWDDNLLRGKPTTHKVFGEVAILSGGALLALAFEHLEAD 205
/ 140 GLTPIAKTRAVSELSTAIAGMGLVGOQKDL-SEGDKPRSA--DAILLTNQFKTSTILFCA 196
b 206 -VSSKKMRAVKELAKSIGTKGLVAGQAKDLSEGLEQNDVGLEDLEYIHVKTGSLEA 264
/ 197 STQWASIAANASCEARENLRHSLDLGQAFOLLDDLTD---GMTDTGKDINQD--AGKST 251
b 265 SAVIGAVGGGTSEKEIKVNFARCIIGLLFQVDDIILDETKSSSEHLGKTAGKQVAGKLT 324
/ 252 LVNLLGSGAVFEELRQHLRLASEHL 276
b 325 YPKVIGVKSKEFVEXLKRDAREHL 349

RESULT 12
87505
eranyltrnstransferase [imported] - Caulobacter crescentus
;Species: Caulobacter crescentus
;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
;Accession: D87505
;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
;B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
;oc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
;Title: Complete Genome Sequence of Caulobacter crescentus.
;Reference number: A87249; MUID:21173698; PMID:11259647
;Accession: D87505
;Status: preliminary
;Molecule type: DNA
;Residues: 1-294 <STO>
;Cross-references: GB:AE005673; NID:g13423548; PIDN:AAK24040.1; GSPDB:GN00148
;Genetics:
;Gene: CC2069

Query Match      23.6%; Score 360.5; DB 2; Length 294;
Best Local Similarity 36.5%; Pred. No. 1.6e-22;
Matches 105; Conservative 39; Mismatches 101; Indels 43; Gaps 11;

/ 21 ADIDS-LRDLQPL-VOGERDCVGAARREGTLAPGKIRPMWLLLTARDLGCASHGGILLD 78
b 12 ADIVTVALDELLPRADGPESRLTEAMRYAALGPKELRPFPALET-----GKMF 61
/ 79 L-----ACAVEMVHAASLIIDMPCWDDAQMRGRPTIHTQYGEHVAILAAVALLSK 130
b 62 LPERPVLRAACALECHIVAYSLVHDDIIPAMVDDVRGRPTVHKQYDEATILAGDALQTA 121

```

Db 223 GALLRFAACEAGAIAGSNQAEQRLRFGEKIGLSFQLADDLLDTADAAATMGKATGKDA 282
QY 248 --GKSTLVNLLGSGAVEERLRQHRLRLASEHLS 277
Db 283 ARGKGTLLVALRGEAWAREKLOQVQVAEASELLA 314
Search completed: February 29, 2004, 14:52:42
Job time : 9.43386 secs

A:Residues: 1-304 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88439.1; PID:g15157936; GSPDB:GN00169
A:Genetics:
A:Gene: AGR_C_4933
A:Map position: circular chromosome
Query Match 23.5%; Score 358.5; DB 2; Length 304;
Best Local Similarity 36.0%; Pred. No. 2.5e-22;
Matches 98; Conservative 44; Mismatches 111; Indels 19; Gaps 7;
17 ECLLADI---DSRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLLLLTARDLGCAISH 73
20 EALLGHLISGEARADEITRPNLLE---NRHGVNLGSKRLRPFLVIESVALLGGDAEA 75
74 GGLLDLACAVEMVHAASLILDDMPQDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFG 133
76 G--LHVGAALCELCYSLVHDDLPAFMDDDLRRGQPTVHRKFDDEATAILAGSLLTLAFD 133
134 VIAEAEGLTPIA---KTRAVSELSTAIGVQGLVQGFQDLSEGDKPRSDAILLTNQPKT 190
134 IIRASDD--NPLAERKALVISLARAGIGMAGGQALDLAAEKAPDEGDIITLQAMKT 191
191 STLFCASTQMASIAANASCEARENLRHRSFLDLGQAFQLDDLTGQMTDT---GKDIHQDA 247
192 GALLRFAACEAGAIAGSNQAEQRLRFGEKIGLSFQLADDLLDTADAAATMGKATGKDA 251
248 --GKSTLVNLLGSGAVEERLRQHRLRLASEHLS 277
252 ARGKGTLLVALRGEAWAREKLOQVQVAEASELLA 283

RESULT 15
AH2910
Genanyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2910
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2910
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAI43702.1; PID:g17741230; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ispA
A:Map position: circular chromosome
C:Superfamily: dimethylallyltransferase
Query Match 23.5%; Score 358.5; DB 2; Length 335;
Best Local Similarity 36.0%; Pred. No. 2.8e-22;
Matches 98; Conservative 44; Mismatches 111; Indels 19; Gaps 7;
17 ECLLADI---DSRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLLLLTARDLGCAISH 73
51 EALLGHLISGEARADEITRPNLLE---NRHGVNLGSKRLRPFLVIESVALLGGDAEA 106
74 GGLLDLACAVEMVHAASLILDDMPQDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFG 133
107 G--LHVGAALCELCYSLVHDDLPAFMDDDLRRGQPTVHRKFDDEATAILAGSLLTLAFD 164
134 VIAEAEGLTPIA---KTRAVSELSTAIGVQGLVQGFQDLSEGDKPRSDAILLTNQPKT 190
165 IIRASDD--NPLAERKALVISLARAGIGMAGGQALDLAAEKAPDEGDIITLQAMKT 222
191 STLFCASTQMASIAANASCEARENLRHRSFLDLGQAFQLDDLTGQMTDT---GKDIHQDA 247

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

% protein - protein search, using sw model

un on: February 29, 2004, 14:27:18 ; Search time 5.10674 Seconds
(without alignments)
3089.496 Million cell updates/sec

itle: US-09-941-947A-26

erfect score: 1526

sequence: 1 LTVCAKKHVLHTGISBQL.....HSTTQLFIQAWFDKKLAWS 303

coring table:

Gapop 10.0 , Gapext 0.5

sarched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Query	Match	Length	DB	ID	Description
1	1344.5	88.1	302	1	CRTE_PANAN			P21684 pantosa ana
2	762.5	50.0	307	1	CRTE_ERWHE			P22873 erwinia her
3	372	24.4	369	1	GGPP_CAPAN			P80042 capsicum an
4	369	24.2	371	1	GGPP_ARATH			P34802 arabidopsis
5	364.5	23.9	366	1	GGPP_SINAL			Q43133 sinapis alb
6	360.5	23.6	297	1	ISPA_BACST			Q08291 bacillus st
7	346.5	22.7	272	1	ISPA_BACSU			P54383 bacillus su
8	345.5	22.6	357	1	GGPP_CATRO			Q4698 caetharanchu
9	332	21.8	299	1	ISPA_SCOLI			P22939 escherichia
10	320	21.0	332	1	ISPA_RHISN			P55339 rhizobium s
11	316	20.7	332	1	ISPA_BRAJA			Q45220 bradyrhizob
12	313.5	20.5	300	1	CRTE_CVAPA			P48368 cyanophora
13	309.5	20.3	295	1	ISPA_HAEIN			P45204 haemophilus
14	307	20.1	288	1	CRTE_RHOSH			P54976 rhodobacter
15	304.5	20.0	291	1	ISPA_MICLU			O65126 micrococcus
16	300	19.7	294	1	ISPA_BUCAP			Q8K9A0 buchnera ap
17	297.5	19.5	289	1	CRTE_RHOCA			P17060 rhodobacter
18	277	18.2	282	1	ISPA_BUCAI			P57537 buchnera ap
19	272	17.8	329	1	ISPB_HAEIN			P44916 haemophilus
20	249.5	16.3	262	1	ISPA_AQUAS			O66952 aquifex aeo
21	246	16.1	334	1	ISDA_METTM			Q53479 methanobact
22	244	16.0	325	1	ISDA_METTH			O26156 methanobact
23	243.5	16.0	323	1	PREA_SVNY3			P72580 synecocyst
24	242	15.9	323	1	ISPB_ECOLI			P19641 escherichia
25	234	15.3	320	1	HEP2_BACST			P55785 bacillus st
26	233.5	15.3	332	1	GGPP_SULSO			P59599 sulfobolus
27	232	15.2	323	1	PREA_CVAPA			P31171 cyanophora
28	231	15.1	359	1	GGPP_WYCTU			O50727 m probable
29	230	15.1	327	1	ISDA_METJA			Q58270 methanococc
30	219.5	14.4	330	1	GGPP_SULAC			P39464 sulfobolus
31	203.5	13.3	348	1	HEP2_BACSU			P31114 bacillus su
32	194.5	12.7	378	1	DPS_SCHPO			O43091 schizosacch
33	193.5	12.7	323	1	PREA_CVACA			Q9C1s1 cyanidium c

ALIGNMENTS

RESULT 1

ID	CRTE_PANAN	STANDARD;	PRT;	302 AA.
AC	P21684;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP synthetase) (Farnesyltransferase).			
GN	CRTE.			
OS	Pantoea ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoea.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
RX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in Escherichia coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
CC	-1- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.			
CC	-1- PATHWAY: Carotenoid biosynthesis.			
CC	-1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
ENBL	D90087; BAAL4124.1;			
PIR	A37802; A37802.			
DR	InterPro: IPR000092; Polyphenyl synt.			
DR	InterPro: IPR008949; Terpenoid synth.			
DR	Pfam: PF00348; polyphenyl synt; 1.			
DR	PROSITE: PS00444; polyphenyl SYNTHET 2; 1.			
DR	PROSITE: PS00723; POLYPHENYL-SYNTHET_1; 1.			
KX	Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.			
SQ	SEQUENCE 302 AA; 32558 MW; CADB04699D2EBA4A CRC64;			
Query Match 88.1%; Score 1344.5; DB 1; Length 302;				
Best Local Similarity 88.4%; Pred. No. 2.6e-101;				
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;				
QY	1 LTVCAKKHVLHTGISBQLIDSLDQLPVGGRDVCVGAARECTLAPGKRIRPMLL 60			
DB	1 MFTVCAKKHVLHTDAABQLADIDRRLDQLLPVEGENDVVGAAAREGALAPGKRIRPMLL 60			
QY	61 LITARDLGCALSHGGLLDLACAVEMVHAASLLIDDMPCMDDAQMRGRPTTIHTQYGEHVA 120			

```

Db 61 LTARDLGCANSHDGLDLACAVKEMVHAASLIIDMPCMDADKRRGRPTIHSYGRHVA 120
QY 121 ILAAVALLSKAFGVIAEAGTPIAKTRAVSELSTA:GMQGLVQGFQKLSGDKPRSD 180
Db 121 ILAAVALLSKAFGVIAEAGTPIAKTRAVSELSTA:GMQGLVQGFQKLSGDKPRSAE 180
QY 181 AILLTNOFKTSTLFCASQOMASIAANASCEARENLRHFSLDLGOAFOLLDDLTGMDTDC 240
Db 191 AILMNHFKTSTLFCASQOMASIVANNASSEARDCLHRSFSLDGOAFOLLDDLTGMDTDC 240
QY 241 KDINQDAGKSTLVNLLSGAVVEERLRQHLRLASHLASAACQNGHSTTQLFIQAWFDKCLA 300
Db 241 KDSNODAGKSTLVNLLSGAVVEERLRQHLRLASHLASAACQNGHSA-TOHFIQAWFDKCLA 299
QY 301 AVS 303
Db 300 AVS 302

RESULT 2
TERTE ERWHE
ID CRUE ERWHE STANDARD; PRT; 307 AA.
AC P22873;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP synthetase) (Farnesyltransferase).
EN CRTE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EH010;
RX MEDLINE=91088634; PubMed=2263648;
RA Armstrong G.A., Albert M., Hearst J.E.;
RT "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).
RN [2]
RP FUNCTION.
RX MEDLINE=92357711; PubMed=1495965;
RA Math S.K., Hearst J.E., Poulter C.D.;
RT "The crtB gene in Erwinia herbicola encodes geranylgeranyl diphosphate synthase.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764(1992).
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + Geranylgeranyl diphosphate.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38424; AAA24819.1; -.
DR EMBL; M87280; AAA64977.1; -.
DR PIR; C39273; C39273.
DR InterPro; IPR000092; Polyrenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF0348; polyrenyl_synth; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
KW Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.
SEQUENCE 307 AA; 33242 MW; 6A534C6194CE9F59 CRC64;

```

```

Query Match 50.0%; Score 762.5; DB 1; Length 307;
Best Local Similarity 56.0%; Pred. No. 2.4e-54;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
QY 17 ROLLADISRLDQLLPVQGERDCVGAARREGTLAPGKRIRPFIHQYGEHVAILAVALSKAFGVIA 136
Db 16 EVMRQSIDDLHAGLLPETDSQDIVSLARREGVMAFGKIRPULLMLAARDLRYQGSMTL 75
QY 77 LDLACAVEMVHAASLIIDMPCMDADKRRGRPTIHSYGRHVAIAVALSKAFGVIA 136
Db 76 LDLACAVELTHTASLMLDMPMDNAELRRGQPTTHKKFGESVAILASVGLSKAFGLIA 135
QY 137 EABGLTPIAKTRAVSELSTA:GMQGLVQGFQKLSGDKPRSDAILTNOFKTSTLFC 196
Db 136 ATGDLPGERRAQAVNELSTAVGQGLVGLQFRLDNDAAALDRTPDAILSTXHLKTGLFSA 195
QY 197 STOMASIAANASCEARENLRHFSLDLGOAFOLLDDLTGMDTDCINQDAGKSTLVNLL 256
Db 196 MLQIVAIASASSPSTRETTLHAFALDQAFQLLDLDRDDHPETGKDRNKDAGKSTLVNRL 255
QY 257 GSGAVEERLRQHLRLASHLASAACQNGHSTTQLFIQAWFDKCLA 300
Db 256 GADAAQKLRHIDSADKHLTFACPGGAIHQ-PHMLWFGHLLA 298

RESULT 3
GGPP CAPAN
ID GGPP CAPAN STANDARD; PRT; 369 AA.
AC P80042;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
EN GGPS1.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Lamuyo; TISSUE=Fruit;
RX MEDLINE=93272043; PubMed=1303794;
RA Kuntz M., Roemer S., Suire C., Huguency P., Weil J.H., Schantz R., Camara B.;
RT "Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening.";
RL Plant J. 2:25-34(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Yolo Wonder;
RX MEDLINE=95195169; PubMed=788631;
RA Badillo A., Steppuhn J., Deruere J., Camara B., Kuntz M.;
RT "Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsicum annuum.";
RL Plant Mol. Biol. 27:425-428(1995).
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans-trans-farnesyl diphosphate.
CC -!- CATALYTIC ACTIVITY: trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key enzyme in plant terpenoid biosynthesis.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

```

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X80267; CAA56554.1; --
PIR; S53722; S53722; Polyprenyl synth.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synth; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
Transferase; Isoprene biosynthesis; Carotenoid biosynthesis;
Chloroplast; Transit peptide. CHLOROPLAST.
CHAIN 1 369
SEQUENCE 369 AA; 40173 MW; 2D527F5B43A29C6F CRC64;
Query Match 24.4%; Score 372; DB 1; Length 369;
Best Local Similarity 34.5%; Pred. No. 9.6e-23;
Matches 99; Conservative 52; Mismatches 108; Indels 28; Gaps 7;
/ 23 IDRLDQLPVQGRDCVGAAMREGLTAPGRKIRPMLLLTARDLGCIAHSGILLDLACA 82
/ 89 VNKALEALIVK-EPHVIHEAMRYSLAGGRVREMLCLAAACELVGG--NQENAAVACA 145
/ 83 VEMVRAASLIDMPCDDAQQVERGRPTHQYGEHVAILAVALLSKAF-GVIAEAGCL 141
/ 146 VEMVHTMSLIDHLPQMDNDLRRGRKTNHLYGDEAVLAGDSLLAFAPFHVNVSTAGV 205
/ 142 TPIAKTRAVSELSTAIGMGLVQGVQFQDLK-EGDKPRSAAILITNPQKTLFCASTOM 200
/ 206 TFSRVGAVAEALAKSIGTEGLVAGQVADIKCTGNASVSLTELFHFVHKTAALLESSVWL 265
/ 201 ASIANASCEARENLRSLDQAPQLDLDTQMTGTGDIQWDAGKSTLVN----- 254
/ 266 GILGGTGTVVEKLRARRARCIGLLFQWDDILD-VTKSELKTAGKADLVVDKTTYPK 324
/ 255 LKSGAVERLRQHLRLASEHLSAACQNGHSTTQLFQWFDKGLAA 301
/ 325 LKLEKAKEFAELNREAKQLEG-----PDSKAA 355
RESULT 4
PFP ARATH STANDARD; PRT; 371 AA.
P34B02; 023201;
01-FEB-1994 (Rel. 28, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP
synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1);
Geranyltransferase (EC 2.5.1.10); Farnesyltransferase
(EC 2.5.1.29)]
GGPSI OR A74G36810.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI TaxID=3702;
[1]
SEQUENCE FROM N.A.
MEDLINE=94286754; PubMed=8016276;
Scolnik P.A., Bartley G.E.;
"Nucleotide sequence of an Arabidopsis cDNA for geranylgeranyl
pyrophosphate synthase."
Plant Physiol. 104:1469-1470(1994).
[2]
SEQUENCE FROM N.A.

```

```

RC STRAIN=cv, Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Sciekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansgore W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen J., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernaier S., Rempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloecher H., Scharfe M., Grimm M., Loehnert T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Agirion A., Vitale D., Liguori R., Piravandi E.,
RA Schnabl S., Hiller R., Schmidt W., Lechman A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacón D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
RA Frishman D., Haase D., Lemcke K., Meves H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan B., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana."
RL Nature 402:769-777(1999).
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of
CC Ipp onto DMAPP to form geranylgeranyl pyrophosphate.
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key
CC enzyme in plant terpenoid biosynthesis.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L25813; AAA32797.1; --
CC EMBL; Z99708; CAB16803.1; --
CC EMBL; AL161590; CAB80347.1; --

```



```

X MEDLINE=95001990; PubMed=7918490;
A Koyana T., Obata S., Saito K., Takeshita-Koike A., Ogura K.;
T "Structural and functional roles of the cysteine residues of Bacillus
T steatotherophilus farnesyl diphosphate synthase.";
L Biochemistry 33:12644-12648(1994).
C !- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
C = diphosphate + trans,trans-farnesyl diphosphate.
C !- SUBCELLULAR LOCATION: Cytoplasmic.
C !- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C
C EMBL: D13293; AAA02551.1; -
R ENBL; S72629; AAB32272.1; -
R ENBL; S72630; AAB32273.2; ALT_SEQ.
R ENBL; S72633; AAB32274.1; -
R ENBL; S72635; AAB32275.2; ALT_SEQ.
R PIR: JX0257; JX0257.
R InterPro: IPR000092; Polyprenyl_synth.
R InterPro: IPR008949; Terpenoid_synth.
R Pfam: PF00348; polyprenyl_synth; 1.
R PROSITE: PS00444; POLYPRENYL SYNTHET 2; 1.
R PROSITE: PS00723; POLYPRENYL SYNTHET 1; 1.
R W TRANSFERASE; Isoprene biosynthesis.
R T MUTAGEN 73 73 C->F,S: NO LOSS OF ACTIVITY.
R T MUTAGEN 289 289 C->F,S: NO LOSS OF ACTIVITY.
R T SEQUENCE 297 AA; 32310 MW; 0F921C3F029EEB6 CRC64;
R
Query Match 23.6%; Score 360.5; DB 1; Length 297;
Best Local Similarity 34.9%; Pred. No. 6.3e-22;
Matches 99; Conservative 41; Mismatches 129; Indels 13; Gaps 5;
Y 14 ISAPOLLADISRLDQLLPQGER-----DCVGNAREGTLAPGKIRPMILLIARDLGC 69
b 4 LSVEQFLNEQKQAVETALSRYIERLEGPAKIKKAWAYSLEAGGKIRPILLUSTVRLGK 63
Y 70 AISHGGILLDACAVEMVHAASLILDMPCMDDAQERGERPTIHTQYGEHVAILAAVALLS 129
b 64 DPAVG--LPVACAIEIMHYSLIHDDLPSEVNDLDRGKPTNKHVFGEMAILAGDGLT 121
Y 130 KAFGVIAEA--EGLTPIAKTRAVSELSTALGMGLVQGFOLSLRGDKRPSADAILITNQ 187
b 122 YAFQLITEIDDERIPPSVRLRIETLRIAKAGPEGMVAGQAADWEGEGKTLTISELEYIHR 181.
Y 186 FKTSTLFCASTOWASTAANASCEARENLRFLSLDGLQAFQLLDLTD---GMTDTGKGIN 244
b 182 HKTGMQLQSVHAGALIGGDARQTELEFPAHILGLAQIRDDILDIEGAEKIGKPVG 241
Y 245 QDA--GKSTLWNLGSGAVEERLRQHLRLASEHLSMAQNG 283
b 242 SDQSNKATVPALLSLAGAKEKLFHIEAQHRLNADVDG 282

```

RESULT 7

SPA_BACSU

D ISPA_BACSU

C P54383;

T 01-OCT-1996 (Rel. 34, Created)

T 01-OCT-1996 (Rel. 34, Last sequence update)

T 10-OCT-2003 (Rel. 42, Last annotation update)

E Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)

E (FPP synthase).

N YQID OR BSU24280.

S Bacillus subtilis.

C Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

X NCBI_TaxID=1423;

N [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
R [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessièr P., Bolotin A., Borchert S.,
RA Borriss R., Bourschi L., Brans A., Braun M., Brignall S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan B., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrero P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Teosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yasane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RT Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D84432; BAA12575.1; -
DR ENBL; Z99116; CAB14359.1; -
DR PIR: A69961; A69961.
DR Subtilist; BG11714; yqid.
DR InterPro: IPR000092; Polyprenyl_synth.
DR InterPro: IPR008949; Terpenoid_synth.
DR Pfam: PF00348; polyprenyl_synth; 1.
DR PROSITE: PS00444; POLYPRENYL SYNTHET 2; 1.
DR PROSITE: PS00723; POLYPRENYL SYNTHET 1; 1.
KW Transferase; Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 272 AA; 29740 MW; 55A26E9A9895D67C CRC64;

```

Query Match

22.7%; Score 346.5; DB 1; Length 272;

Best Local Similarity 34.0%; Pred. No. 7.6e-21;

Matches 91; Conservative 52; Mismatches 108; Indels 17; Gaps 7;

-1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; D00694; BAA00599.1; -
 EMBL; AE000148; AAC73524.1; -
 EMBL; U82664; AAB40177.1; -
 PIR; JQ0665; JQ0665.
 SWISS-2DPAGE; P22939; COLI.
 EcoGene; E010508; ispa.
 InterPro; IPR000092; Polyrenyl_synth.
 InterPro; IPR008949; Terpenoid_synth.
 Pfam; PF00348; polyrenyl_synth_1.
 PROSITE; PS00444; POLYPRENYL_SYNTHET_1; 1.
 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 TRANSFERASE; Isoprene biosynthesis; Complete proteome.
 SEQUENCE 299 AA; 32159 MW; 15BAD05E135060CA CRC64;
 Query Match 21.8%; Score 332; DB 1; Length 299;
 Best Local Similarity 33.8%; Pred. No. 1.3e-19;
 Matches 101; Conservative 46; Mismatches 124; Indels 28; Gaps 8;
 1 LTVCAKHEVHLTGISAEQLADISRLDQLLPVOCERDCVGAAMREGTAPGKIRPMLL 60
 7 LEACVQK-----ANQAL-----SRFAPLPFQNT--VVEIWOYGALLGKRLPFLV 52
 61 LITARDLCAISHGGLLDLACAVENWHAASLIDMPCKDDQMRGRPTIHTOYGEHVA 120
 53 YATGHMFG--VSTNTLDAPAAVECTHAYSILHDDLPAMDDDLRLRGILPTCHVKFGEANA 110
 121 ILAAVALLSKAFGVIAEGLTPIAKTR---AVSELSTAIGMGLVQGOFGDLSEGDKP 176
 111 ILUGDALOTLAFSLISDAD--MPEVSDRDRISMISELASAGIAGMCGGQALDLDAGKH 168
 177 RSADAILLNTQKSTLTFCASTQMASIAA--NASCARENLHRFSLDGLQAPOLLDDLTGD 235
 169 VPLDALERHHRKHTGALIRAAVRGALSAGDKGRALPVLQDYAESIGLAFVQVQDILDV 228
 236 MDTG-----XDINDACKSTLVNLGSGAVEERLRQLRLRLASEHLAAQNGHSTQL 289
 229 VGDITLGRQAGDQQLGKSTYPALLGLRQARKARDLIDDRSLKQLAEQSLDTSAL 287
 RESULT 10
 ID ISPA RHISN STANDARD; PRT; 332 AA.
 AC 255339;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable geranyltranstransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase) (FPP synthase).
 GN Y4KU.
 OS Rhizobium sp. (strain NGR234).
 PE Plasmid sym pNGR234a.
 PC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 PC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 NCBI_TaxID=394;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=97305956; PubMed=9163424;
 FREIBERG C.A., FELLAY R., BAIRDOCH A., BROUGHTON W.J., ROSENTHAL A., PERRET X.;
 "Molecular basis of symbiosis between Rhizobium and legumes.";
 Nature 387:394-401 (1997).
 -1- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.

-1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 -1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; AE000082; AAB91752.1; -
 PIR; T10875; T10875.
 InterPro; IPR000092; Polyrenyl_synth.
 InterPro; IPR008949; Terpenoid_synth.
 Pfam; PF00348; polyrenyl_synth_1.
 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 TRANSFERASE; Isoprene biosynthesis; Plasmid.
 SEQUENCE 332 AA; 34688 MW; EE68C1547449AA87 CRC64;
 Query Match 21.0%; Score 320; DB 1; Length 332;
 Best Local Similarity 35.4%; Pred. No. 1.3e-18;
 Matches 102; Conservative 40; Mismatches 106; Indels 40; Gaps 10;
 23 IDSRLDQLLPVQ--GERDCVGAAMREGTAPGKIRPMLLITARDLCAISHGGLLDLA 80
 44 VEQALRLLCADHEDGTELM--AAMRYATLHGKGRTR--ALLCIAAGALADTFAP--ILDVVG 100
 81 CAVENWHAASLIDMPCKDDQMRGRPTIHTOYGEHVAITLAAVALLSKAFGVIA--EA 138
 101 ALEEMWHAETLVHDDLPAWDDVLRGLPTVHVKEGENTAILVGDALQAHAFLLTASLDA 160
 139 EGLTPIAKTRAVSELSTAIGMGLVQGOFGDLSEGDKPRSADAILLNTQKSTLTFCAST 198
 161 PDENRIALVR---ELAAVSAEAGAGGQAMDLSLVGKVELDRIVAMHRMCKGALVRASV 217
 199 QWASIAAANASCARENH---RFSLDLQAPOLLDDLTGQWTD---GKINDOAG--K 249
 218 RMGALCAEAEDADATLYCALDHYSAFCGLALQVDDILDATATLTGKTPGKDAQAQK 277
 250 STLNVLLSGAVE-----ERLRQHLRLASEHL 276
 278 PTCASIMGLQAARQFALDLCAGEATAIPLGPRAEARLAQMLQASAVL 325
 RESULT 11
 ID ISPA BRAJA STANDARD; PRT; 332 AA.
 AC Q45220;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable geranyltranstransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase) (FPP synthase).
 GN FPPS OR BLR2148.
 OS Bradyrhizobium japonicum.
 PC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 PC Bradyrhizobiaceae; Bradyrhizobium.
 NCBI_TaxID=375;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=USDA 110;
 TULLY R.E., KEISTER D.L.;
 "Cloning and mutagenesis of a cytochrome P-450 locus from Bradyrhizobium japonicum that is expressed anaerobically and symbiotically.";
 Appl. Environ. Microbiol. 59:4136-4142 (1993).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=USDA 110;
 MEDLINE=98322110; PubMed=9655913;
 TULLY R.E., VAN BERKUM P., LOVINS K.W., KEISTER D.L.;

RT "Identification and sequencing of a cytochrome P450 gene cluster from
 RT *Bradyrhizobium japonicum*".
 RL Biochim. Biophys. Acta 1398:243-255 (1998).
 RW [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 UA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 UA Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,
 UA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 UA Tabata S.;
 XT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 XT *Bradyrhizobium japonicum* USDA110.";
 RL DNA Res. 9:189-197 (2002).
 IC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 IC = diphosphate + trans-trans-farnesyl diphosphate.
 IC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 IC -!- SIMILARITY: Belongs to the PPP/GGPP synthetase family.
 IC
 IC This SWISS-PROT entry is copyright. It is produced through a collaboration
 IC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 IC the European Bioinformatics Institute. There are no restrictions on its
 IC use by non-profit institutions as long as its content is in no way
 IC modified and this statement is not removed. Usage by and for commercial
 IC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 IC or send an email to license@isb-sib.ch).
 IC
 IC EMBL; U12678; AAC28894.1;
 IC EMBL; AP005942; BAC47413.1; ALT_INIT.
 IC PIR; I40213; I40213.
 IC InterPro; IPR000092; Polyrenyl_synth.
 IC InterPro; IPR008949; Terpenoid_synth.
 IC Pfam; PF00348; Polyrenyl_synth; 1.
 IC PROSITE; PS00444; POLYPRENYL_SYNTHET 2; 1.
 IC PROSITE; PS00723; POLYPRENYL_SYNTHET 1; 1.
 IC Transferrase; Isoprene biosynthesis; Complete proteome.
 IC SEQUENCE 332 AA; 34638 MW; 47644FAF0220CE4C CRC64;

Query Match 20.7%; Score 316; DB 1; Length 332;
 Best Local Similarity 33.4%; Pred. No. 2.9e-18;
 Matches 107; Conservative 45; Mismatches 128; Indels 40; Gaps 11;

2Y 8 HVHILGTISAEQLLA-----DIDSLDQLLPVQ--GERDVCGAAM 44
 2E 8 HDDRAGVSANGILQAQAGSGLLPPIWQDGAKRVEALRLCAEDDGETELM-AAM 66
 2Y 45 REGTLAPGKRIRPMLLLTARDLCAISHGGLDLACAVENWHAASLIIDMPCMDAQM 104
 2B 67 RYATLHGGRTR-ALLCLAGALADTPAH-MLDDVGAAIEMHACTLVHDDLPAMDVYL 124
 2Y 105 RRGRTTHQYGERVAAILAAVALLSKAPGVIA--EAGLTPIAKTRAVSELSTAIGWQGL 162
 2B 125 RRGLTPTVHVKEGATILVGDALQAHPLTLASINAFSDSPIALVR---ELAQVSAEGA 181
 2Y 163 VQGQFKDLSEGDKPRSDAAILLTNQFKTSTLFCASTOMASIAA---NASCEARE-NLHRP 218
 2B 182 AGGQAIDLVLGKGVHEDRIVAMHRMKSGLVSRVSMGALCAVGNAAHAALYCALDHY 241
 2Y 219 SLDLQAFQLDLDLTGMDTDT--GKDINODAG--KSTIANLLGSGAVEERLQHLRLAS 273
 2B 242 SACTGLAQVDDLDLVTAUTAAAGKTPGKDNAAQKPTCASIMGLQBAROFALDLDLRDAG 301
 2Y 274 EHLASAQNGHSTTQLFTQA 293
 2B 302 EALAPLGPRAERLAQLIORA 321

RESULT 12
 CITE CYAPA
 ID CITE CYAPA STANDARD; PRT; 300 AA.
 AC P48368;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase homolog (EC 2.5.1.29) (GGPP
 DE synthetase) [Farnesyltransferase].
 DE CITE.
 OS Cyanophora paradoxa.
 OG Cyanelle.
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 XT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
 RL Plant Mol. Biol. Rep. 13:327-332 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 XT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
 XT the genetic complexity of a primitive plastid.";
 RL (In) Schenk H.B.A., Herrmann R., Jeon K.W., Mueller N.E.,
 RL Schwemmler W. (eds.);
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).
 CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 CC -!- SIMILARITY: Belongs to the PPP/GGPP synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U30821; AAA81312.1; --
 CC PIR; U06969; T06969.
 CC InterPro; IPR000092; Polyrenyl_synth.
 CC InterPro; IPR008949; Terpenoid_synth.
 CC Pfam; PF00348; Polyrenyl_synth; 1.
 CC PROSITE; PS00444; POLYPRENYL_SYNTHET 2; 1.
 CC PROSITE; PS00723; POLYPRENYL_SYNTHET 1; 1.
 CC Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 CC Isoprene biosynthesis; Transferrase; Cyanelle.
 CC SEQUENCE 300 AA; 33008 MW; 05950D0E3BA04934 CRC64;
 Query Match 20.5%; Score 313.5; DB 1; Length 300;
 Best Local Similarity 33.2%; Pred. No. 3.9e-18;
 Matches 87; Conservative 54; Mismatches 102; Indels 19; Gaps 8;
 QY 23 IDSLDQLLPVQGRDVCGAAMREGTLAPGKRIRPMLLLTARDLCAISHGGL---LDL 79
 DB 21 VEDTLNKSIP-RGNPTFIYDSIRYLSAGGKRIRPILCLAS-----CELAGGTMEVALPT 74
 QY 80 ACAYEMVHAASLIIDMPCMDAQMREGRTTHQYGERVAAILAAVALLSKAPGVIA-EA 138
 DB 75 ACALEMTHMSLIHDDLPAMDNDSYRKGKTNHIIYGEDLAILAGDALLAYAFEFIATQT 134
 QY 139 EGLTPIAKTRAVSELSTAIGWQGLVQGFQKDL-SEGDKPRSDAAILLTNQFKTSTLFCAS 197
 DB 135 KNVPADLTVKVIQAQVAHSVITTSGLVGGQIIDLSESGKSDTTLFTLNFIHTKTKALLEAA 194
 QY 198 TQMASIAANASCEARENLRPSLDLQAFQLDLDLTGMDT---DTGKDINOD--AGKSTL 252
 DB 195 VLGGALLAGAKEMDNRFYRQAQNIQAFQLIIDDVLDDIISTEEKLGSIGKDLTKTKATY 254
 QY 253 VNLGSGAVEERLQHLRLASE 274

```
255 PSFVG---VERSIOKQAEELVEE 273
      ISPA_HAEIN STANDARD; PRT; 295 AA.
P45204;
01-NOV-1995 (Rel. 32, Created)
01-NOV-1995 (Rel. 32, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
(ZPP synthase).
159A OR H11438;
Haemophilus influenzae.
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
STRAIN-Rd / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7342800;
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Uttarback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
"Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd";
Science 269:496-512(1995).
-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
= diphosphate + trans,trans-farnesyl diphosphate.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; U02822; AAC23087.1; -.
FIR; C64123; C64123.
TIGR; H11438; -.
InterPro; IPR000092; Polyprenyl synth.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synth; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
W TRANSFERASE; Isoprene biosynthesis; Complete proteome.
Q SEQUENCE 295 AA; 32407 MW; A64DB850266E670F CRC64;
Query Match 20.3%; Score 309.5; DB 1; Length 295;
Best Local Similarity 34.1%; Pred. No. 8.1e-18;
Matches 88; Conservative 46; Mismatches 107; Indels 17; Gaps 7;
Y 20 LADISRLDQLLPVQGE-----RDCVGAMREBGTAPGKRIPMLLLLTARDLQCAISHG 74
b 8 LQOVQTRINFRLEAFQEGIESHNAPLEAMKYALLGGKVRPFLWYATGQMLG---AEK 64
Y 75 GLLD-LACAVENMVAASLIDMPQCMDDAQMRGRPTHTQVGHVATIAAVALLSKAPG 133
D 65 QTLDYAAAEALHAYSLHDDLPAWDDNLERRGHTCHIQFDEATALLAGDALQSAPE 124
Y 134 VTAESGLPIPAKTRAVSELSAIGXQGVHQQQFQNDLSEGDKPSADAILTNQFKSTL 193
D 125 ILTKPTNISTEOKLAIQILAQAGAGVQGNCLGQSLDLSSEHKQISLSSELELHAKTGAL 184
QY 194 FCASTOMASIANASCEAR--ENLHRFSLDGLQAFQLLDLTD---GMTDTGKDI--NQD 246
Db 185 LIAALKLGFICSPHFTDKRLQSLTQYAEALGLAFQVQDDILDIEGDSAEITGKQVGADLD 244
QY 247 ACKSTLVNLG-SGAVER 263
Db 245 LDKSTYPKLLGLSGAKQK 262
RESULT 14
CRTE RHOSH STANDARD; PRT; 288 AA.
AC P54976; Q9RFC5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP
DE synthetase) (farnesyltransferase).
GN CRTE.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=95338278; PubMed=7721699;
Lang H.F., Cogdell R.J., Takaichi S., Hunter C.N.;
RT "Complete DNA sequence, specific tms insertion map, and gene
RT assignment of the carotenoid biosynthesis pathway of Rhodobacter
RT sphaeroides.";
RL J. Bacteriol. 177:2064-2073(1995).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=2015911; PubMed=10648776;
Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1.";
RL Nucleic Acids Res. 28:862-867(2000).
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AJ010302; CAB38744.1; -.
EMBL; AF195122; AAF24294.1; -.
FIR; S49625; S49625.
FIR; T50750; T50750.
InterPro; IPR000092; Polyprenyl synth.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synth; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Isoprene biosynthesis; Transferase.
FT CONFLICT 53 53 A -> R (IN REF. 1).
FT CONFLICT 57 57 D -> S (IN REF. 1).
FT CONFLICT 60 60 V -> A (IN REF. 1).
FT CONFLICT 123 123 R -> G (IN REF. 1).
FT CONFLICT 128 128 Q -> R (IN REF. 1).
FT CONFLICT 234 234 N -> S (IN REF. 1).
FT CONFLICT 247 247 R -> G (IN REF. 1).
FT CONFLICT 264 264 G -> A (IN REF. 1).
FT CONFLICT 277 277 E -> D (IN REF. 1).
```

SQ SEQUENCE 288 AA; 30256 MW; 5A7BDEBC41811A36 CRC64;
 Query Match 20.1%; Score 307; DB 1; Length 288;
 Best Local Similarity 35.5%; Pred. No. 1.3e-17;
 Matches 88; Conservative 32; Mismatches 96; Indels 32; Gaps 8;
 43 ANRGDTLAP-----GRIIPMLLLTARDLGCAISHGGLLDLAC-AVEKWH 87
 16 ARGQSEAPSKLATALDYAVTPGGRIIRPTLLSVA--TACGDDRPVLSDAVALELIH 73
 88 AASLILDDMPCCMDDAQMERGRPTHTQYGEHVAILAAVALLSKAFGVIAEAGLTPIAKT 147
 74 CASLHVHDLPCFDDAEIRRGRTVRAISEPLAHLTGDSLWVGFEVLARAADQFORAL 133
 148 RAVSELSTAIGM-QGLVGOQFKDSEGDKPRSDAILLTNQFKTSTLFCASTOMASTARN 206
 134 QLVTALAVRTGMPMGICAGQ-----GWSESQINLSAYHRAKTGALFIATQMGATAAG 187
 207 ASCEARENLRHRSFLDLGQAFOLLDTGMDT---GKIDNQDA--GKSTIVNLGSGAV 261
 188 YEAFWEEL---GARIGAFQVADDLRLDALCAETLKGPAQODEIHARPNVREYGVEGA 244
 262 EERLROHL 269
 245 AKRLKDL 252
 RESULT 15
 ISPA_MICLU STANDARD; PRT; 291 AA.
 AC 066126;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
 DE (GPP synthase).
 OS Micrococcus luteus (Micrococcus lysodeikticus).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococcineae; Micrococcaceae; Micrococcus.
 OX NCBI_TaxID=1270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B-P 26;
 RX MEDLINE=98175686; PubMed=9515931;
 RA Shimizu N., Koyama T., Ogura K.;
 RT "Molecular cloning, expression, and characterization of the genes
 encoding the two essential protein components of Micrococcus luteus
 B-P 26 hexaprenyl diphosphate synthase.";
 RL J. Bacteriol. 180:1578-1581(1998).
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 = diphosphate + trans,trans-farnesyl diphosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -!- SIMILARITY: Belongs to the GPP/GGPP synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB003187; BAA25265.1; -;
 DR InterPro; IPR000092; Polyrenyl_synth.
 DR InterPro; IPR008949; Terpenoid synth.
 DR Pfam; PF00348; polyrenyl synth. 1.
 DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
 DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
 KW Transferase, isoprene biosynthesis.
 SQ SEQUENCE 291 AA; 32362 MW; 797125AB71E5674A CRC64;
 Query Match 20.0%; Score 304.5; DB 1; Length 291;

Best Local Similarity 36.2%; Pred. No. 2e-17;
 Matches 92; Conservative 36; Mismatches 109; Indels 17; Gaps 7;
 QY 24 DSRDLQLLPVGERDCVGAAMREGTLAPGRIIPMLLLTARDLGCAISHGGLLDLACAV 83
 DB 18 ESLNRYHPAGSR---LHEAINYSLSAGGRIIRPULVLTUDDSLG-GNAHDG-LPFGIAL 72
 QY 84 EMVHAASULDDMPCCMDDAQMERGRPTHTQYGEHVAILAAVALLSKAFGVIAEAGLTP 143
 DB 73 EMHTYSLIHDDLPAMDNDYRRGKLTNHRKFDDEATAILAGDALLTDAFQCILNQ-LNA 131
 QY 144 IAKTRAVSELSTAIGM-QGLVGOQFKDSEGDKPRSDAILLTNQFKTSTLFCASTOMASI 203
 DB 132 ETKLSLIMLLSTASGSGNMVYQGLDMQGEHKTUTLNELERIHIHKGELTRAAIVSAGI 191
 QY 204 AANASCEARENLRHRSFLDLGQAFOLLDTGMDT---GMDTGTGKIDNQAGKSTLVNL 255
 DB 192 INMFDAQIEQLNIIGKVGMLMFQIKDDILDVEGSPENIKGT-VGSDLNND--KSTTVSL 248
 QY 255 LGSGAVEERLROHL 269
 DB 249 LGLRASKOLLNDKL 262
 Search completed: February 29, 2004, 14:45:03
 Job time : 6.10674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on on: February 29, 2004, 14:51:24 ; Search time 20.1175 Seconds
(without alignments)

3180.293 Million cell updates/sec

File: US-09-941-947a-26

irect score: 1526

quence: 1 LTVCARKHVLGTISABQLL.....HSTTLQFIQAFDKLAAYS 303

oring table: BLZSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 21153259 residues

al number of hits satisfying chosen parameters: 809742

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1526	100.0	303	10	US-09-941-947a-26
2	1526	100.0	303	14	US-10-218-118-2
3	427	28.0	295	14	US-10-166-225A-185
4	425	27.9	295	9	US-09-547-267-1
5	425	27.9	295	10	US-09-920-923-2
6	412.5	27.0	294	15	US-10-369-493-17358
7	394.5	25.9	292	15	US-10-369-493-10617
8	388	25.4	269	13	US-10-108-915-18
9	381.5	25.0	294	15	US-10-369-493-18010
10	380.5	24.9	291	15	US-10-369-493-8506
11	376.5	24.7	287	9	US-09-925-637-64
12	376.5	24.7	287	14	US-10-084-205-64
13	375.5	24.6	293	9	US-09-815-242-12583
14	375.5	24.6	293	10	US-09-920-923-37
15	368.5	24.1	288	9	US-09-815-242-5239

Sequence 16943, A
Sequence 12014, A
Sequence 19922, A
Sequence 11449, A
Sequence 14816, A
Sequence 15003, A
Sequence 2608, Ap
Sequence 10187, A
Sequence 18752, A
Sequence 2, Appli
Sequence 20893, A
Sequence 23201, A
Sequence 4670, Ap
Sequence 7429, Ap
Sequence 45, Appl
Sequence 16549, A
Sequence 22, Appl
Sequence 14, Appl
Sequence 72, Appl
Sequence 20, Appl
Sequence 10069, A
Sequence 732, App
Sequence 15817, A
Sequence 16200, A
Sequence 15449, A
Sequence 15191, A
Sequence 9546, Ap
Sequence 14084, A
Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-941-947a-26
; Sequence 26, Application US/09941947a
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picstagio, Steve
; APPLICANT: Rouverie, Pierre B.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 26
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947a-26

Query Match 100.0%; Score 1526; DB 10; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.3e-150;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LTVCARKHVLGTISABQLLADISRLDQLLPVQGRDCVGAAMEGTLAPKRIKRPMLL 60
Db 1 LTVCARKHVLGTISABQLLADISRLDQLLPVQGRDCVGAAMEGTLAPKRIKRPMLL 60
Qy 61 LITARDLGCAISHGGLLDLACAVEMVHAASLILDDMPCVDDAQMGRGRTTHTQYGEHVA 120

db 61 LLTARDLGCAISHGGLDLACAVEMVHAASLIIDMPCDDAQRRGRPTIHTQYGERVA 120
yy 121 ILAAVALLSKAFGVIABEGTLPIAKTRAVSELSTALGMQGLVQGFQKDLSEGDKERSAD 180
db 121 ILAAVALLSKAFGVIABEGTLPIAKTRAVSELSTALGMQGLVQGFQKDLSEGDKERSAD 180
yy 181 AILLTNQFKSTLFCASTQMASIAANASCBARENHRFSLDLGQAFOLLDDLTGDMTDTG 240
db 181 AILLTNQFKSTLFCASTQMASIAANASCBARENHRFSLDLGQAFOLLDDLTGDMTDTG 240
yy 241 KQINQDAGKSTLVNLLGSGAVEERLROHLRLASEHLSAACQNGHSTTQLPFIQAWFDKKLA 300
db 241 KQINQDAGKSTLVNLLGSGAVEERLROHLRLASEHLSAACQNGHSTTQLPFIQAWFDKKLA 300
yy 301 AVS 303
db 301 AVS 303

RESULT 2

JS-10-218-118-2
Query Match 100.0%; Score 1526; DB 14; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.3e-150;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
db 1 LTVCAKKEVHLTGISABOLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
db 1 LTVCAKKEVHLTGISABOLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
yy 61 LLTARDLGCAISHGGLDLACAVEMVHAASLIIDMPCDDAQRRGRPTIHTQYGERVA 120
db 61 LLTARDLGCAISHGGLDLACAVEMVHAASLIIDMPCDDAQRRGRPTIHTQYGERVA 120
yy 121 ILAAVALLSKAFGVIABEGTLPIAKTRAVSELSTALGMQGLVQGFQKDLSEGDKERSAD 180
db 121 ILAAVALLSKAFGVIABEGTLPIAKTRAVSELSTALGMQGLVQGFQKDLSEGDKERSAD 180
yy 181 AILLTNQFKSTLFCASTQMASIAANASCBARENHRFSLDLGQAFOLLDDLTGDMTDTG 240
db 181 AILLTNQFKSTLFCASTQMASIAANASCBARENHRFSLDLGQAFOLLDDLTGDMTDTG 240
yy 241 KQINQDAGKSTLVNLLGSGAVEERLROHLRLASEHLSAACQNGHSTTQLPFIQAWFDKKLA 300
db 241 KQINQDAGKSTLVNLLGSGAVEERLROHLRLASEHLSAACQNGHSTTQLPFIQAWFDKKLA 300
yy 301 AVS 303
db 301 AVS 303

RESULT 3

US-10-166-225A-185
; Sequence 185, Application US/10166225A
; Publication No. US20030148416A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMBELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISEEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/10/166.225A
; CURRENT FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 185
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Paracoccus sp. RI534
US-10-166-225A-185
Query Match 28.0%; Score 427; DB 14; Length 295;
Best Local Similarity 39.6%; Pred. No. 6.8e-36;
Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
QY 23 IDSRLDQLLPVQGERDCV---GAAMREGTLAPGKRIRPMLLLLTARDLGCAISHG--- 75
DB 13 VEIRLAQ---ISGQGVVSAPLGAANSDAALSPGKRFRAVLMLWAE-----SSGGVCD 63
QY 76 -LLDLACAVEMVHAASLIIDMPCDDAQRRGRPTIHTQYGERVAIIAAVALLSKAFGV 134
DB 64 AMVDAACAVEMVHAASLIIDMPCDDARTERGQPATHVAHGGRVAVLAGIALITAMRI 123
QY 135 IAAEGLTPIAKTRAVSELSTALGMQGLVQGFQKDLSEGDKERSADAILLTNOFKSTLEF 194
DB 124 LGHAGATPDQARLVAASBRAMPVGLCAGQDLDL---HAPKDAAGIBREQDLKTGVLP 180
QY 195 CASTQMASIAANASCBARENHRFSLDLGQAFOLLDDLTGDMTDTG---TGKQINQD-AGKS 250
DB 181 VAGLEMLSIITKGLDKAETEQLMAPGRLGVRVFSYDDLLDVIQDKASTGKTGRTDAAFG 240
QY 251 TLVNLGSGAVEERLROHLRLASEHLSAACQNGHSTTQLP 290
DB 241 PIRGLMAVGQMGD-VAOHYRASRAQLDELMR-----TRLF 274
RESULT 4
US-09-547-267-1
; Sequence 1, Application US/09547267
; Patent No. US20020147371A1
; GENERAL INFORMATION:
; APPLICANT: Hohmann, Hans-Peter
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/547,267
; FILING DATE:

APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10617
LENGTH: 292
TYPE: PRT
ORGANISM: Spingomonas aromaticivorans
US-10-369-493-10617

Query Match 25.9%; Score 394.5; DB 15; Length 292;
Best Local Similarity 36.0%; Pred. No. 1.6e-32;
Matches 95; Conservative 45; Mismatches 115; Indels 9; Gaps 5;

22 DIDRLDQLLPVQGE-RDCVGAAMREGTLAPGKRIKRPMLLLLTARDLGCAISHGGLDLA 80
11 DIDRSFLLLPVQGPDRDLVEMRYAAIGGKRLRLVLCATGALHG--VSRRAVRVG 68
81 CAVEVMVHAASLLDDMPDMDAQMRRGPTIHTQYGEHVAILAAVALLSKAFGVIAE-AE 139
69 TAIEAHVYSLIHDLPDMDNDGLRHGKPTVHAFDDATAVLAGDALLHDFAPVLSDPAT 128
140 GLTPIAKTRAVSELSTAIGMGLVQGVQKDL-SEGDKPRSDAAILLTNQFTSTLFCASTQ 199
129 SGDFPTRIELTIRLTATAGMNGMCGGQMDIVATSSFDLPVTRLOQLKGTGALLGASVE 188
200 MASTAANASCEARENLERFSLDLCQAFOLLDDLTGMDT---TGKIDNQA--GKSTLVN 254
189 LGAILGKVPPEGRTHLRGARDIGLAFQIADDLDAEGDEAAGKALKRDAAGKETFLS 248
255 LLGSGAVEERLRQHLRLASEHLSA 278
249 LLGADRAEQARLLVNQAVEHLAS 272

RESULT 6
US-10-108-915-18
; Sequence 18, Application US/10108915
; Publication No. US20020177204A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Shen, Jennie
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
; FILE REFERENCE: BB1286 US NA
; CURRENT APPLICATION NUMBER: US/10/108,915
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Glycine max
US-10-108-915-18

Query Match 25.4%; Score 388; DB 13; Length 369;
Best Local Similarity 37.3%; Pred. No. 1.1e-31;
Matches 98; Conservative 49; Mismatches 104; Indels 12; Gaps 6;

23 IDSRDLQLLPVQGERDCVGAAMREGTLAPGKRIKRPMLLLLTARDLGCAISHGGLDLACA 82

Db 89 VNOALDAIAIALRDPHK-IHQAMEYSLLAGGKRVKPVLCIAACELVGG--TEATAIPAACA 145
Qy 83 VEVYHAASLLDDMPDMDAQMRRGPTIHTQYGEHVAILAAVALLSKAF-GVIAEAEG 141
Db 146 VEMHTWLSLIHDLPDMDNDLGRGKPTNKHVGVAVLAGDALLAPAFEHVAASLEGV 205
Qy 142 TPIAKTRAVSELSTAIGMGLVQGVQKDL-SEGDKPRSDAAILLTNQFTSTLFCASTQ 200
Db 206 SPSRVRAIGELAKSIGTEGLVAGQVVDIDSEGVANVGLTLEFIHVHKTAAALLEAAV 265
Qy 201 ASIAANASCEARENLERFSLDLCQAFOLLDDLTGMDTGTGKIDNODAGKS-----TLVN 254
Db 266 GAIVGGSDVEVEKLRKPARCIGLLFQVDDIILD-VTKSEELGKTAGKDLVADKVTYPK 324
Qy 255 LLGSGAVEERLRQHLRLASEHLS 277
Db 325 LLGIDKSKEFAQELLKDAKEOLS 347

RESULT 9
US-10-369-493-18010
; Sequence 18010, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18010
; LENGTH: 294
; TYPE: PRT
; ORGANISM: SPHINGOMONAS
US-10-369-493-18010

Query Match 25.0%; Score 381.5; DB 15; Length 294;
Best Local Similarity 36.9%; Pred. No. 3.7e-31;
Matches 101; Conservative 42; Mismatches 110; Indels 21; Gaps 9;

Qy 18 QULADIDRLDQLLPVQGE-RDCVGAAMREGTLAPGKRIKRPMLLLLTAR----DLGCAIS 72
Db 10 QVAAEIDDQFDRLLEVPDDPRADLYRAMHAAIGGGKRLRPLLVFATAQLFSVTKSCAR 69
Qy 73 HGGLLDLACAVEMVHAASLLDDMPDMDAQMRRGPTIHTQYGEHVAILAAVALLSKAF 132
Db 70 -----VATALECHIVSVLSLIHDDLPAWDDDDMRGKPTVHKAFDEATAILAGDCLHDLAF 123
Qy 133 GVIA-EAEBGLTPIAKTRAVSELSTAIGMGLVQGVQKDLSEGDKPRSDAAILLT--NQPK 189
Db 124 ELIAHEATHADPFVRVELISELARSSGPAKGAGQMDL-EAEK-ASFDLATVTRLOALK 181
Qy 190 TSTLFCASTQWASTAANASCEARENLERFSLDLCQAFOLLDDLTGMDT---TGKIDNOD 246
Db 182 TGALESCAVECGAILGRVPHEGRTGLGRTADIGLAFQIVDDILDAGSDEAVVKGKLGKD 241
Qy 247 --AGKSTLVNLLGSGAVEERLRQHLRLASEHLSA 278
Db 242 GAAGKETFLSLIGIDRAREQARMLVDQAIHLHA 275

RESULT 10
US-10-369-493-8506
; Sequence 8506, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xiaofeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 8506
 LENGTH: 291
 TYPE: PRT
 ORGANISM: Ralstonia metallidurans
 3-10-363-493-8506

Query Match 24.9%; Score 380.5; DB 15; Length 291;
 Best Local Similarity 39.6%; Pred. No. 4.6e-31;
 Matches 93; Conservative 34; Mismatches 91; Indels 17; Gaps 4;

43 ANREGLAPGRIRPMILLITARDLG-----CAISHGGLLDLACAVEMVHAASLILDDMP 97
 35 AMRYAVLGGKRRPRLVHAAGEVVGATPEACDAA-----ACAVEMIHAYSLSVHDMP 87
 98 CNDMDQMERGRPTHTQYGEHVAILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAI 157
 88 CNDMDLARGPTVEKAYDEATALLVGDALQSQAFLVLAQTPALAAEARLKVVAELAVAS 147
 158 GMOGLVQGFQDLSEGDPRSDAILLTNQKSTLFCASTOMASIAANASCEARENLR 217
 148 GSGMCGQQAIDLQNVGKAMTREALGHRMKTGALLRASVRMGALCGNIDQAGLVALDR 207
 218 FSLDLGQAFQDLDDLTGMDTD---GKINODAG--KSTLVNLLSGGAVBELRQ 267
 208 YAAVGLAFQVDDILDVTRVATLGTAGKDAANDKFTYVSLGLDAARELAQA 262

RESULT 11
 3-09-925-637-64
 Sequence 64, Application US/09925637
 Patent No. US2002010338A1
 GENERAL INFORMATION:
 APPLICANT: Choi
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides
 FILE REFERENCE: PB560
 CURRENT APPLICATION NUMBER: US/09/925,637
 CURRENT FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: PCT/US00/23773
 PRIOR FILING DATE: 2000-08-31
 PRIOR APPLICATION NUMBER: US 60/151,933
 PRIOR FILING DATE: 1999-09-01
 PRIOR APPLICATION NUMBER: US 08/781,986
 PRIOR FILING DATE: 1997-01-03
 PRIOR APPLICATION NUMBER: US 08/956,171
 PRIOR FILING DATE: 1997-10-20
 PRIOR APPLICATION NUMBER: US 60/009,861
 PRIOR FILING DATE: 1996-01-06
 NUMBER OF SEQ ID NOS: 74
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 64
 LENGTH: 287
 TYPE: PRT
 ORGANISM: Homo sapiens
 3-09-925-637-64

Query Match 24.7%; Score 376.5; DB 9; Length 287;
 Best Local Similarity 35.5%; Pred. No. 1.2e-30;
 Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5;

11 LTGISAEQLLADISRLDQLLPVQGERDCVGAAMREGTL----APGKIRPMILLITARD 66

Db 1 MTNLPNKLIDEVNNELS-----VAINKSVMDTQLEESMLYSLNAGGKIRPVLILLTLD 56
 QY 67 LGCAISHGGLLDLACAVEMVHAASLILDDMPQMDDAQMRGRPTHTQYGEHVAILA 126
 Db 57 LNTVEYELG--MKSALALEMIHTYSLIHDDLPAMDNDYRRGKLTNHKVVGEWTAILAGDA 114
 QY 127 LLSKAPGVIAEAGLTPIAKTRAVSELSTAIGMOGLVQGFQDLSEGDKPRSADAILTN 186
 Db 115 LUTKAPELISSDDRLTDEVKIKVLQRLSIASGHVGMVGGQMLDMQSEGQPIDLETLEMIH 174
 QY 187 QKSTLFCASTOMASIAANASCEARENLRPSLDLGOAFOLLDDLTGMDTD---TGKDI 243
 Db 175 KTKTGALLTFAVMSAADIANVDDITTKHELESYSYHLGVMFQKDDLLDCYGEAKLGKKV 234
 QY 244 NQDA--GKSTLVNLLSGGAVBELRQHLRLASEHLS 277
 Db 235 GSDLENNKSTYVSLGKGDAEDKLTYYHDAVADEL 270

RESULT 12
 US-10-084-205-64
 Sequence 64, Application US/10084205
 Publication No. US20030049648A1
 GENERAL INFORMATION:
 APPLICANT: Choi, Gil
 TITLE OF INVENTION: 37 Staphylococcus aureus Genes and Polypeptides
 FILE REFERENCE: PB15P1
 CURRENT APPLICATION NUMBER: US/10/084,205
 CURRENT FILING DATE: 2002-02-28
 PRIOR APPLICATION NUMBER: PCT/US00/23773
 PRIOR FILING DATE: 2000-08-31
 PRIOR APPLICATION NUMBER: 60/151,933
 PRIOR FILING DATE: 1999-09-01
 NUMBER OF SEQ ID NOS: 74
 SOFTWARE: PatentIn Ver. 3.1
 SEQ ID NO 64
 LENGTH: 287
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-10-084-205-64

Query Match 24.7%; Score 376.5; DB 14; Length 287;
 Best Local Similarity 35.5%; Pred. No. 1.2e-30;
 Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5;

QY 11 LTGISAEQLLADISRLDQLLPVQGERDCVGAAMREGTL----APGKIRPMILLITARD 66
 Db 1 MTNLPNKLIDEVNNELS-----VAINKSVMDTQLEESMLYSLNAGGKIRPVLILLTLD 56
 QY 67 LGCAISHGGLLDLACAVEMVHAASLILDDMPQMDDAQMRGRPTHTQYGEHVAILA 126
 Db 57 LNTVEYELG--MKSALALEMIHTYSLIHDDLPAMDNDYRRGKLTNHKVVGEWTAILAGDA 114
 QY 127 LLSKAPGVIAEAGLTPIAKTRAVSELSTAIGMOGLVQGFQDLSEGDKPRSADAILTN 186
 Db 115 LUTKAPELISSDDRLTDEVKIKVLQRLSIASGHVGMVGGQMLDMQSEGQPIDLETLEMIH 174
 QY 187 QKSTLFCASTOMASIAANASCEARENLRPSLDLGOAFOLLDDLTGMDTD---TGKDI 243
 Db 175 KTKTGALLTFAVMSAADIANVDDITTKHELESYSYHLGVMFQKDDLLDCYGEAKLGKKV 234
 QY 244 NQDA--GKSTLVNLLSGGAVBELRQHLRLASEHLS 277
 Db 235 GSDLENNKSTYVSLGKGDAEDKLTYYHDAVADEL 270

RESULT 13
 US-09-815-242-12583
 Sequence 12583, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12583
LENGTH: 293
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12583
Query Match 24.6%; Score 375.5; DB 9; Length 293;
Best Local Similarity 35.6%; Pred. No. 1.6e-30;
Matches 98; Conservative 43; Mismatches 119; Indels 15; Gaps 5;
Yy 11 LTGSAQLADISRLDQLLPVQGERDCVGAANRGTL-----APKRIKRPMLLITARD 66
Yb 1 MTNLPNNKLIDVNNLS-----VAINKSVMDTQLEBSMLYSLNAGGKIRPVLILLTLD 56
Yy 67 LGCAISHGLDLACAVEMVHAASLIIDDDMPCHDDAOMRRGRPTIHTQVGEHVAILAAVA 126
Yb 57 LNTYEYLG--KMSAIALEMHTYSLIHDDLPAMDNDYRGKLTNKHVYGEWTAIAGDA 114
Yy 127 LLSKAPGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAILLTN 186
Yb 115 LTKAFELISSDRLTDEVKIKVLQRLSIASGHVGMVGGQMLDMQSEGQPIDLTLETMT 174
Yy 187 QFKTSTLFCASTOMASIAANASCEARENHRSFLDLGQAFQLDDLTGQMTD---TGKDI 243
Yb 175 KTKGALLTFVWGAAADIANVDDATKGLHLSYVHLGWMFQIKDDLLDCYGEAKLGKV 234
Yy 244 NQDA--GKSTLVNLGSGAVEERLROHLRLASEHL 276
Yb 235 GSDLENNKSTVSLGKGABDKLTYHRDAVDL 269
RESULT 14
US-09-920-923-37
Sequence 37, Application US/09920923
Publication No. US2003002273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 56

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 293
TYPE: PRT
ORGANISM: E-396
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Unknown
US-09-920-923-37
Query Match 24.6%; Score 375.5; DB 10; Length 293;
Best Local Similarity 38.3%; Pred. No. 1.6e-30;
Matches 90; Conservative 38; Mismatches 86; Indels 21; Gaps 6;
Yy 23 IDSLDQLLPVQ---ERDCVGAAMREGLAPKRIKRPMLLITARDLGCALSHGG---- 75
Yb 13 LQTLLEI--AQGFVAVSQPLGPMGSHGALSGLKFRGMLMLAAE-----ASGGVCDT 64
Yy 76 LLDLACAVEMVHAASLIIDDDMPCHDDAOMRRGRPTIHTQVGEHVAILAAVALLSKAPGVI 135
Yb 65 IVDACAVEMVHAASLIIDDDMPCHDDAOMRRGRPTIHTQVGEHVAILAAVALLSKAPGVI 124
Yy 136 ABAEGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAILLTNQPKTSTLFC 195
Yb 125 AGARGASGTVRAQLVRLSSRLGPOGLCAGQDLDLHAA---KNGAGVBOEDLKTGVLFI 181
Yy 196 ASTOMASIAANASCEARENHRSFLDLGQAFQLDDLTGQMTD---TGKDIHQDA 247
Yb 182 AGLEMLAVIKFPAEAGTQMIDFGRQLGRVFSYDLDLVGDDAALGKDTGDA 236
RESULT 15
US-09-815-242-5239
Sequence 5239, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5239
LENGTH: 288
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5239
Query Match 24.1%; Score 368.5; DB 9; Length 288;
Best Local Similarity 36.2%; Pred. No. 8.1e-30;
Matches 97; Conservative 41; Mismatches 115; Indels 15; Gaps 5;

18 QLLADISRLDQLLEVOGERDCVGAAMEGTL-----APGKRIKIRPMLLLLTARDIGCAISH 73
3 KLIDEVNNELS-----VAINKSVMDTQLEESMLYSLNAGGKIRPVLTLTLDLSLNTYEL 58
74 GCLLDLACAVENWHAASLIIDDMPCMDDAQMERGRPTHTQYGEHVALLAAVALLSKAFG 133
59 G--MKSAIALEMIHTYSLIHDDLPAMDNDYRGRKLTNHKVIKGTWALLAGDALLTKAFE 116
134 VIAEAGLTPIAKTRAVSELSTAIGMOGLVOGQFMDLSEGDKPRSDAAILITNOFKTSTL 193
117 LISSDDRLTDEVKIKVIGLQSLIASGHVGVGGQMLDMQSEGQPIDLETLEMIHKTGTAL 176
194 PCASTOMASIAANASCBARENLRFSLDLGOAFQLDDLTGMDT---TGKDINQDA--G 248
177 LTFVMSAADTIANVDDATKEHLESYSYHLGMVFOIKDDLLDCYGDCAKLGKKVGSLENN 236
249 KSTLVNLLGSGAVEERLRQHLRLASEHL 276
237 KSTYVSLLGKGAEDKLTYHRDAVDEL 264

Search completed: February 29, 2004, 15:27:54
>b time : 22.1175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 9.20761 Seconds
(without alignments)
1698.885 Million cell updates/sec

title: US-09-941-947A-26

effect score: 1526

sequence: 1 LTVCKKHVHTGISAEQLL.....HSTQLFIQWFDKLAAYS 303

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/PTCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1344.5	88.1	302	1 US-07-783-705A-1	Sequence 1, Appli
2	763.5	50.0	298	1 US-08-095-726-4	Sequence 4, Appli
3	763.5	50.0	298	1 US-08-096-043-4	Sequence 4, Appli
4	763.5	50.0	298	1 US-08-093-577-4	Sequence 4, Appli
5	763.5	50.0	298	1 US-08-096-623A-4	Sequence 4, Appli
6	763.5	50.0	307	1 US-08-095-726-2	Sequence 2, Appli
7	763.5	50.0	307	1 US-08-096-043-2	Sequence 2, Appli
8	763.5	50.0	307	1 US-08-093-577-2	Sequence 2, Appli
9	763.5	50.0	307	1 US-08-096-623A-2	Sequence 2, Appli
10	425	27.9	295	3 US-08-660-645A-1	Sequence 1, Appli
11	425	27.9	295	3 US-09-298-718-1	Sequence 1, Appli
12	425	27.9	295	3 US-09-546-969-1	Sequence 1, Appli
13	425	27.9	295	3 US-08-980-832-2	Sequence 2, Appli
14	425	27.9	295	4 US-09-547-267-1	Sequence 2, Appli
15	425	27.9	295	4 US-09-920-923B-2	Sequence 2, Appli
16	375.5	24.6	293	3 US-08-980-832-37	Sequence 37, Appli
17	375.5	24.6	293	3 US-09-320-923B-37	Sequence 37, Appli
18	367.5	24.1	293	3 US-09-276-873-2	Sequence 2, Appli
19	360.5	23.6	297	1 US-08-534-910B-8	Sequence 8, Appli
20	360.5	23.6	297	3 US-08-886-466-2	Sequence 2, Appli
21	360.5	23.6	297	3 US-09-475-304-2	Sequence 2, Appli
22	360.5	23.6	297	4 US-09-101-126-3	Sequence 3, Appli
23	360.5	23.6	297	4 US-09-367-528A-5	Sequence 5, Appli
24	359.5	23.6	297	1 US-08-534-910B-7	Sequence 7, Appli
25	359.5	23.6	297	1 US-08-534-910B-9	Sequence 9, Appli
26	358.5	23.5	287	1 US-08-534-910B-10	Sequence 10, Appli
27	357.5	23.4	285	3 US-09-187-050-12	Sequence 12, Appli

Sequence 2, Appli
Sequence 14, Appli
Sequence 16, Appli
Sequence 18, Appli
Sequence 20, Appli
Sequence 22, Appli
Sequence 24, Appli
Sequence 26, Appli
Sequence 27, Appli
Sequence 28, Appli
Sequence 29, Appli
Sequence 30, Appli
Sequence 31, Appli
Sequence 32, Appli
Sequence 33, Appli
Sequence 34, Appli
Sequence 1, Appli
Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-07-783-705A-1
; Sequence 1, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 13-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-1

Query Match 88.1%; Score 1344.5; DB 1; Length 302;

Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIDSRLDQLPVQGRDVCVGAAMREGTLPAGKIRPMLLLLTARDLCAISHGGL 76
7 EVMRQSIDHLAGLLPETSQDIVSLAMREGVWAPGKIRPMLMLLAARDLRYQGSMPYL 66
77 LDLCACAVEMVHAASLILDDMPQDDAOMRGRPTIHTQYGEHVAILAVALLSKAPGVIA 136
67 LDLCACAVELTHTASLMLDDMPQDDAOMRGRPTIHTQYGEHVAILAVALLSKAPGVIA 126
137 EASGLTPIAKTRAVSELSTAIGMQGLVQGOQKDLSEGDKPRSDAILLTNQFKTSLFCA 196
127 ATGDLPGERRAQAVNELSTAVGLQGLVQGFRLDNDALDRTPDAILSTNHLKTGILFSA 186
197 STQMASTAAANASCEARENLRHFSLDLQAPQLDLDLTGDMTDTGKDINDAGKSTLVNLL 256
187 MLQIVAIASASSPSTRETLHAFALDQAPQLDLDLDDHPETGDKORNDACKSTLVNRL 246
257 GSGAVERLRBOHLRLASEHLSAACQNGHSTTQLFQIOWFDKILA 300
247 GADARQKUSEHIDSADKHLTFACPOGGAIRO-FMHLWFGHILA 289

```

Sequence 4, Application US/08093577
Patent No. 5545816

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukhartji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Phycoene Biosynthesis in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,577
FILING DATE: 19-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,569
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5545816val B
TELEPHONE: 3128567180
TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

-08-093-577-4

Query Match 50.0%; Score 763.5; DB 1; Length 298;
Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

. QY 17 EQLADIDSRLDQLPVQGRDVCVGAAMREGTLPAGKIRPMLLLLTARDLCAISHGGL 76
DB 7 EVMRQSIDHLAGLLPETSQDIVSLAMREGVWAPGKIRPMLMLLAARDLRYQGSMPYL 66
QY 77 LDLCACAVEMVHAASLILDDMPQDDAOMRGRPTIHTQYGEHVAILAVALLSKAPGVIA 136
DB 67 LDLCACAVELTHTASLMLDDMPQDDAOMRGRPTIHTQYGEHVAILAVALLSKAPGVIA 126
QY 137 EASGLTPIAKTRAVSELSTAIGMQGLVQGOQKDLSEGDKPRSDAILLTNQFKTSLFCA 196
DB 127 ATGDLPGERRAQAVNELSTAVGLQGLVQGFRLDNDALDRTPDAILSTNHLKTGILFSA 186
QY 197 STQMASTAAANASCEARENLRHFSLDLQAPQLDLDLTGDMTDTGKDINDAGKSTLVNLL 256
DB 187 MLQIVAIASASSPSTRETLHAFALDQAPQLDLDLDDHPETGDKORNDACKSTLVNRL 246
QY 257 GSGAVERLRBOHLRLASEHLSAACQNGHSTTQLFQIOWFDKILA 300
DB 247 GADARQKUSEHIDSADKHLTFACPOGGAIRO-FMHLWFGHILA 289

```

RESULT 5

US-08-096-623A-4
Sequence 4, Application US/08096623A
Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukhartji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500

TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 amino acids
TYPE: amino acid
TOPOLOGY: linear
STRANDEDNESS:
MOLECULE TYPE: protein
US-08-095-726-4

Query Match 50.0%; Score 763.5; DB 1; Length 298;
Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
Y 17 EQLLADIDSRDLQVQGERDCVGAAMEGRTAPGKIRIRPMLLLLTARDLGCALSHGGL 76
b 7 EVNRQSIDHLAGLLPETUSQDIVSLAMEGVMAFGKIRIRPLMLLLAARDLRYQSGMPTL 66
Y 77 LDACAVEMVHAASLILDDMPQVDDQMRGRPTTHQYGEHVAILAAVALLSKAFGVIA 136
b 67 LDACAVELTHTASLMLDDMPQVDDQMRGRPTTHQYGEHVAILAAVALLSKAFGVIA 126
Y 137 EASGLTPIAKTRAVSELSTAIGMVGQVQFQDLSEGDKPRSDAILLTNOKFTSLFCA 196
b 127 ATGDLPCERRAQAVALNELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTLGILFSA 186
Y 197 STQVASTAANASCEARENLRHRSFLDLGQAFQQLDDLTGQMTDTGKDIQNDACKSTLVNLL 256
b 187 MQAIVAIASASSPSTRETLHAFALDFQAFQQLDDLRDDHPETGKRNKDGKSTLVNRL 246
Y 257 GSCAVERLRQHLRLASEHLSAACQNGHSTTQFQAWFQKLA 300
b 247 GADAAARQKLEHIDSADKHLTFACPOGGAIRQ-FMHLWFGHLLA 289

ESULT 6
S-08-095-726-2
Sequence 2, Application US/08095726
Patent No. 5530188
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530188val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
Best Local Similarity 56.0%; Pred. No. 1.6e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
QY 17 EQLLADIDSRDLQVQGERDCVGAAMEGRTAPGKIRIRPMLLLLTARDLGCALSHGGL 76
DB 16 EVNRQSIDHLAGLLPETUSQDIVSLAMEGVMAFGKIRIRPLMLLLAARDLRYQSGMPTL 75
QY 77 LDACAVEMVHAASLILDDMPQVDDQMRGRPTTHQYGEHVAILAAVALLSKAFGVIA 136
DB 76 LDACAVELTHTASLMLDDMPQVDDQMRGRPTTHQYGEHVAILAAVALLSKAFGVIA 135
QY 137 EASGLTPIAKTRAVSELSTAIGMVGQVQFQDLSEGDKPRSDAILLTNOKFTSLFCA 196
DB 136 ATGDLPCERRAQAVALNELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTLGILFSA 195
QY 197 STQVASTAANASCEARENLRHRSFLDLGQAFQQLDDLTGQMTDTGKDIQNDACKSTLVNLL 256
DB 196 MQAIVAIASASSPSTRETLHAFALDFQAFQQLDDLRDDHPETGKRNKDGKSTLVNRL 255
QY 257 GSCAVERLRQHLRLASEHLSAACQNGHSTTQFQAWFQKLA 300
DB 256 GADAAARQKLEHIDSADKHLTFACPOGGAIRQ-FMHLWFGHLLA 298

RESULT 7
US-08-096-043-2
Sequence 2, Application US/08096043
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in
Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 307 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 3-08-096-043-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
 Best Local Similarity 56.0%; Pred. No. 3.6e-79;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

17 EQLADIISRLDQLPVQGERDCVGAAMREGTLPAGKRIIRPMLLLTARDLGCAISHGGL 76
 16 EVMROSIDHLAGLIPETDSQDIVSLAWRGVMAFGKRIIRPMLLLAARDLRYQGSMTL 75
 77 LDLACAVEMVHAASILDDMPGMDAOMRGEPHTHTYQGEHVAIILAAVALLSKAFGVIA 136
 76 LDLACAVELTHTASLMLDDMPGMDNAELRRGQPTTHKKGESVAILASVGLLSKAFGLIA 135
 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFQDLSEGDKPRASDAIILTNQFTSTLFCA 196
 136 ATGDLPGERRAQNVELSTAVGLQGLVGLQFEDLNDALDRTPDAILSTNHLKTGILFSA 195
 197 STOMASIAANASCEARENHRSFSLDLGOAFQLLDLDLTDGMDTDTGKDINQDAGKSTLVNLL 256
 196 MLQIVAIASASSPSTRETLHAFALDFGOAFQLLDLDLDDHPETGKRNKDKAGKSTLVNRL 255
 257 GSGAVEERLROHLRLASHELSAACQNGHSTTOLFQAWFDKCLA 300
 256 GADAAQKRLREHIDSADKELTPACPGGAIRQ-FMHLWFGHLLA 298

RESULT 8

3-08-093-577-2

Sequence 2, Application US/08093577

Patent No. 5545816

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.

APPLICANT: Brinkhaus, Friedhelm L.

APPLICANT: Mukharji, Indrani

APPLICANT: Proffitt, John H.

APPLICANT: Yarger, James G.

APPLICANT: Yen, Hwei-Che B.

TITLE OF INVENTION: Phytoene Biosynthesis in

TITLE OF INVENTION: Genetically Engineered Hosts

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Amoco Corp., Patents and Licensing Dept

STREET: 200 E Randolph St

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/093,577

FILING DATE: 19-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/785,569

FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5545816val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3128567180

TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 307 amino acids

TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-093-577-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
 Best Local Similarity 56.0%; Pred. No. 1.6e-79;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

17 EQLADIISRLDQLPVQGERDCVGAAMREGTLPAGKRIIRPMLLLTARDLGCAISHGGL 76
 16 EVMROSIDHLAGLIPETDSQDIVSLAWRGVMAFGKRIIRPMLLLAARDLRYQGSMTL 75
 77 LDLACAVEMVHAASILDDMPGMDAOMRGEPHTHTYQGEHVAIILAAVALLSKAFGVIA 136
 76 LDLACAVELTHTASLMLDDMPGMDNAELRRGQPTTHKKGESVAILASVGLLSKAFGLIA 135
 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFQDLSEGDKPRASDAIILTNQFTSTLFCA 196
 136 ATGDLPGERRAQNVELSTAVGLQGLVGLQFEDLNDALDRTPDAILSTNHLKTGILFSA 195
 197 STOMASIAANASCEARENHRSFSLDLGOAFQLLDLDLTDGMDTDTGKDINQDAGKSTLVNLL 256
 196 MLQIVAIASASSPSTRETLHAFALDFGOAFQLLDLDLDDHPETGKRNKDKAGKSTLVNRL 255
 257 GSGAVEERLROHLRLASHELSAACQNGHSTTOLFQAWFDKCLA 300
 256 GADAAQKRLREHIDSADKELTPACPGGAIRQ-FMHLWFGHLLA 298

RESULT 9

US-08-096-623A-2

Sequence 2, Application US/08096623A

Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.

APPLICANT: Brinkhaus, Friedhelm L.

APPLICANT: Mukharji, Indrani

APPLICANT: Proffitt, John H.

APPLICANT: Yarger, James G.

APPLICANT: Yen, Hwei-Che B.

TITLE OF INVENTION: Biosynthesis of Zeaxanthin and

TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts

NUMBER OF SEQUENCES: 104

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Welsh & Katz, Ltd.

STREET: 120 S. Riverside Plaza, 22nd Floor

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096,623A

FILING DATE: 22-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/805,061

FILING DATE: 09-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/662,921

FILING DATE: 28-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/562,674

FILING DATE: 03-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/525,551

FILING DATE: 18-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/487,613

FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Camson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
Best Local Similarity 56.0%; Pred. No. 1.6e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

QY 17 EQLADIISRLDQLLPVQGERDCVGNREGLTAPGKRIKIRPMLLLLTARDLGAISHGGL 76
DB 16 EVMQSIDHLAGLLPETDSQDIIVSLAMREGVMAGKRIKIRPMLLLLTARDLGAISHGGL 75
QY 77 LLDLACAVEMVHAASLILDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAFGVIA 136
DB 76 LLDLACAVELTHASLMLDDMFCMDNAELRRGQPTTHKKFGESVAILASVGLLSKAFGLIA 135
QY 137 EAGLTPTIATKTRAVSELSTAIGMQLVQGFQKLSGDKPRASDAIILLTNQKTTSLP 194
DB 136 ATGDLPGERRAQAVNELSTAVGLQSLVQGFQKLSGDKPRASDAIILLTNQKTTSLP 195
QY 197 STOMASIAANASCEARENLERHSLDGLQAFOLLDDLTGMDTGTGKINODAGKSTLWNL 256
DB 196 MLIQVAILASASPSTETILHAFALDFGQAFOLLDDLRDDHETGKRNKAGKSTLWNL 255
QY 257 GSGAVVEERLQHLRLASEHLSAACQNGHSTTOLFQIAWFDKKLA 300
DB 256 GADAAQKLRHIDSGAKHLTFACPGGAIRQ-FMELWFGHLLA 298

RESULT 10
US-08-660-645A-1
Sequence 1, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-660-645A-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;
Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
QY 23 IDSRLDQLLPVQGERDCV---GAAMEEGTLAPGKRIKIRPMLLLLTARDLGAISHGGL 75
DB 13 VEIRLAQ---ISGQGVVSAPLGAASDAALSFGKRFRAVLMMLVAE-----SSGGVCD 63
QY 76 -LLDLACAVEMVHAASLILDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAPGV 134
DB 64 AKVDAACAVEMVHAASLILDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAPGV 123
QY 135 IARABGLTPIATKTRAVSELSTAIGMQLVQGFQKLSGDKPRASDAIILLTNQKTTSLP 194
DB 124 LGARAGATPDQARILVASMSRAMPVGLCAGDLDL---HAPKDAAGIEREQDLKTGVLF 180
QY 195 CASTOMASIAANASCEARENLERHSLDGLQAFOLLDDLTGMDTGTGKINOD-AGKS 250
DB 181 VAGLEMLSIIRKGLDKAETEQLKAFGRQGRVFGVYDLDLVIGDKASTGKXTARDTAPG 240
QY 251 TLVNLGSGAVVEERLQHLRLASEHLSAACQNGHSTTOLF 290
DB 241 PKGGLMAYQMGD-VAQHYRASRAQLDELME-----TFLP 274

RESULT 11
US-09-298-718-1
Sequence 1, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

-09-546-969-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFLGAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTTHQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAIGMOGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLP 180

195 CASTOMASIAANASCEARENLRHFSLDLGOAQLDLDTGQMTD---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETQLMAFGRLGRVFSYDLDLDVIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVYERLRHLRLASEHLSAAQNGHSTTOLF 290

241 PKGGLMVGQMGD-VACHYRASRAQLDELNR-----TRLF 274

SUET 12

-09-546-969-1

Sequence 1, Application US/09546969

Patent No. 6207409

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSES: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nucleu

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/546,969

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-546-969-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFLGAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTTHQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAIGMOGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLP 180

195 CASTOMASIAANASCEARENLRHFSLDLGOAQLDLDTGQMTD---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETQLMAFGRLGRVFSYDLDLDVIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVYERLRHLRLASEHLSAAQNGHSTTOLF 290

241 PKGGLMVGQMGD-VACHYRASRAQLDELNR-----TRLF 274

RESULT 13

US-08-980-832-2

Sequence 2, Application US/08980832B

Patent No. 6291204

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/08/980,832B

CURRENT FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 2

LENGTH: 295

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

US-08-980-832-2

Query Match

27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFLGAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTTHQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAIGMOGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLP 180

195 CASTOMASIAANASCEARENLRHFSLDLGOAQLDLDTGQMTD---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFSYDDLLDVIGDKASTGKTARDTAAPG 240
251 TLVNLGSGAVERLRQHLRLASEHLSAAQONCHSTTOLF 230
241 PKGLMAVGQMGD-VAGHYRASRAQLDELNR-----TRLF 274

RESULT 14

US-09-547-267-1

Sequence 1, Application US/09547267

Patent No. 6613543

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Icom, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/547,267

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-547-267-1

Query Match

Best Local Similarity 27.9%; Score 425; DB 4; Length 295;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSRLDQLLPVQGERDCV-----GAAMREGTLAPKGRIRPMLLLLTARDLGCALSHGG--- 75

13 VEIRLAQ---ISQFGVVSAPLGAMSDAALSFGKFRVAVLMVAE-----SSGGVCD 63

76 -LLDLACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 123

135 IAAEGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPESADAILLTNQFKTSTLF 194

124 IGEARGATPDQARLVASMRAMPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLF 180

195 CASTOMASIAANASCEARENHLRFLSLDGOAFQLDLDTGMDT---TGKINDQ-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFSYDDLLDVIGDKASTGKTARDTAAPG 240

QY 251 TLVNLGSGAVERLRQHLRLASEHLSAAQONCHSTTOLF 290
DB 241 PKGLMAVGQMGD-VAGHYRASRAQLDELNR-----TRLF 274

RESULT 15

US-09-920-923B-2

Sequence 2, Application US/09920923B

Patent No. 6677134

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Fermentative Carotenoid Production

FILE REFERENCE: 15464 US (C38435/125944)

CURRENT APPLICATION NUMBER: US/09/920,923B

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832

PRIOR FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 295

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

US-09-920-923B-2

Query Match

Best Local Similarity 27.9%; Score 425; DB 4; Length 295;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSRLDQLLPVQGERDCV-----GAAMREGTLAPKGRIRPMLLLLTARDLGCALSHGG--- 75

13 VEIRLAQ---ISQFGVVSAPLGAMSDAALSFGKFRVAVLMVAE-----SSGGVCD 63

76 -LLDLACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVEMVHAASLIIDDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 123

135 IAAEGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPESADAILLTNQFKTSTLF 194

124 IGEARGATPDQARLVASMRAMPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLF 180

195 CASTOMASIAANASCEARENHLRFLSLDGOAFQLDLDTGMDT---TGKINDQ-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFSYDDLLDVIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVERLRQHLRLASEHLSAAQONCHSTTOLF 290

241 PKGLMAVGQMGD-VAGHYRASRAQLDELNR-----TRLF 274

Search completed: February 29, 2004, 14:54:55

Job time : 11.2076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: February 29, 2004, 14:26:38 ; Search time 35.8246 Seconds
(without alignments)
2389.754 Million cell updates/sec

title: US-09-941-947A-26

effect score: 1526

sequence: 1 LTVCAKHVHLTGISAEQLL.....HSTTQLFIQAWFDKLAAYS 303

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atbase :

A Genesep29Jan04:*
1: Genesep1980s:*
2: Genesep1980s:*
3: Genesep1980s:*
4: Genesep2000s:*
5: Genesep2000s:*
6: Genesep2003as:*
7: Genesep2003bs:*
8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1526	100.0	303	5	Aae22311 Pantoea s
2	1526	100.0	303	6	Abp96685 Pantoea s
3	1518	99.5	303	6	Aaol6020 Pantoea s
4	1344.5	88.1	302	2	Aar07463 Polypepti
5	1344.5	88.1	302	2	Aaw82255 C. utilis
6	1344.5	88.1	302	2	Aaw87888 Protein e
7	1344.5	88.1	302	2	Aaw99098 Erwinia u
8	763.5	50.0	298	2	Aaw01120 Recombina
9	763.5	50.0	298	2	Aaw00171 N termina
10	763.5	50.0	298	2	Aaw00342 N-termina
11	763.5	50.0	298	2	Aaw32470 Erwinia h
12	763.5	50.0	307	2	Aaw01119 Geranylge
13	763.5	50.0	307	2	Aaw00170 Geranylge
14	763.5	50.0	307	2	Aaw00341 Wild type
15	763.5	50.0	307	2	Aaw32469 Geranylge
16	762.5	50.0	298	2	Aar13982 Geranylge
17	761.5	49.9	307	2	Aar13981 Geranylge
18	643	42.1	332	6	Abm70119 Photorhab
19	427	28.0	295	6	Abu97246 Enzyme po
20	425	27.9	295	2	Aaw06515 Flavobact
21	425	27.9	295	2	Aaw69530 Flavobact
22	425	27.9	295	7	ADB74152 Flavobact
23	388	25.4	369	6	Aae25932 Soybean G
24	388	25.4	369	6	Abu67329 Soybean G
25	377.5	24.7	347	5	ABB91794 Herbicida

ALIGNMENTS

RESULT 1

Aae22311
ID AAE22311 standard; protein; 303 AA.

XX
AC AAE22311;

XX
DT 25-JUL-2002 (first entry)

DE Pantoea stewartii geranylgeranyl pyrophosphate synthase (CrTE) enzyme.

XX
KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;
KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KW aquaculture; enzyme; geranylgeranyl pyrophosphate synthase; CrTE.

XX
OS Pantoea stewartii.

XX
FN WO200218617-A2.

XX
PD 07-MAR-2002.

XX
PF 04-SEP-2001; 2001WO-US027420.

XX
PR 01-SEP-2000; 2000US-0229858P.

XX
PR 01-SEP-2000; 2000US-0229907P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
Odom JM, Picataggio SK, Rouviere PB;

WPI; 2002-351711/38.

DR N-PSDB; AAD35509.

XX
PT Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by
PT using microorganisms having a nucleic acid molecule encoding enzymes in
PT the carotenoid biosynthetic pathway and which metabolize single carbon
PT substrates.

XX
PS Claim 15; Page 134-135; 156pp; English.

XX
CC The invention relates to a method for producing carotenoid compounds. The
CC method comprises a transformed metabolising host cell, comprising
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
CC control of regulatory sequences, and contacting the host cell with carbon
CC substrate to produce a carotenoid compound. The method is useful for
CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by
CC using microorganism having a nucleic acid molecule encoding enzymes in

Aau00859 S. aureus
Abm71047 Staphyloc
Abb92479 Herbicida
Aaw69537 Flavobact
Aau36990 Staphyloc
Abu16439 Protein e
Aag42844 Arabidops
Aag42843 Arabidops
Abp99379 Arabidops
Abb93265 Herbicida
Aau33743 Staphyloc
Abb92493 Herbicida
Aab23333 Staphyloc
Aag07001 Arabidops
Aag42713 Arabidops
Aag42712 Arabidops
Aag07000 Arabidops
Aag42711 Arabidops
Aag06999 Arabidops
Abb91864 Herbicida

26 376.5 24.7 287 4 AAU00859
27 376.5 24.7 293 6 ABM71047
28 376.5 24.7 357 5 ABB92479
29 375.5 24.6 293 2 AAW69537
30 375.5 24.6 293 4 AAU36990
31 375.5 24.6 293 6 ABU16439
32 369 24.2 295 3 AAG42844
33 369 24.2 371 3 AAG42843
34 369 24.2 371 5 ABP99379
35 369 24.2 371 5 ABB93265
36 368.5 24.1 288 4 AAU33743
37 368 24.1 360 5 ABB92493
38 367.5 24.1 293 3 AAB23333
39 364 23.9 287 3 AAG07001
40 364 23.9 287 3 AAG42713
41 364 23.9 288 3 AAG42712
42 364 23.9 288 3 AAG07000
43 364 23.9 376 3 AAG42711
44 364 23.9 376 3 AAG06999
45 364 23.9 376 5 ABB91864

CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Pantoea stewartii geranylgeranyl pyrophosphate synthase (CrtE) enzyme
 CC used in the invention

XX SQ Sequence 303 AA;

Query Match 100.0%; Score 1526; DB 5; Length 303;
 Best Local Similarity 100.0%; Pred. No. 2.6e-152;
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2Y 1 LTVCAKXHVHTGTISAEQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 DB 1 LTVCAKXHVHTGTISAEQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 2Y 61 LITARDLGCALSHGGLLDLACAVEMVHAASLLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 DB 61 LITARDLGCALSHGGLLDLACAVEMVHAASLLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 2Y 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSD 180
 DB 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSD 180
 2Y 181 AILLTNQPKTSTLFCASTQMASIAANASCEARENLRHRSFLDLGQAFOLLDDLTGMDTGT 240
 DB 181 AILLTNQPKTSTLFCASTQMASIAANASCEARENLRHRSFLDLGQAFOLLDDLTGMDTGT 240
 2Y 241 KDINQDACKSTLVNLLSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIOWFQKLA 300
 DB 241 KDINQDACKSTLVNLLSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIOWFQKLA 300
 2Y 301 AVS 303
 DB 301 AVS 303

RESULT 2

ABP96685
 ID ABP96685 standard; protein; 303 AA.

XX AC ABP96685;

XX DT 03-JUN-2003 (first entry)

XX DE Pantoea stewartii geranylgeranyl pyrophosphate synthase SEQ ID NO:2.

XX KW Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
 KW crtI; crtB; crtZ; geranylgeranyl pyrophosphate synthase; enzyme;
 KW phytoene; carotenoid.

XX OS Pantoea stewartii.

XX PN WO2003016503-A2.

XX PD 27-FEB-2003.

XX PF 15-AUG-2002; 2002WO-US026647.

XX PR 15-AUG-2001; 2001US-0312646P.

XX PA (DUPO) DU PONT DE NEMOURS & CO B I.

XX PI Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

XX DR WPI; 2003-268323/26.

XX DR N-PSDB; ACC44759.

XX Novel nucleic acid molecule isolated from Pantoea stewartii encoding a
 PT carotenoid biosynthetic enzyme, useful for regulating carotenoid
 PT biosynthesis in an organism.

XX PS Claim 4; Page 57-58; 68pp; English.

XX The present invention describes Pantoea stewartii carotenoid biosynthetic
 CC enzymes (I). More specifically described are the geranylgeranyl
 CC pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
 CC lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
 CC (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to
 CC ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
 CC carotenoid biosynthesis in an organism, by over-expressing (I) in an
 CC organism, such that the carotenoid biosynthesis is altered in the
 CC phytoene to the carotenoids, for creating recombinant organisms that have
 CC the ability to produce various carotenoid compounds, and also for
 CC enhancing or manipulating carotenoid compounds. (I) can also be used for
 CC producing gene products having enhanced or altered activity

XX SQ Sequence 303 AA;

Query Match 100.0%; Score 1526; DB 6; Length 303;
 Best Local Similarity 100.0%; Pred. No. 2.6e-152;
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LTVCAKXHVHTGTISAEQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 DB 1 LTVCAKXHVHTGTISAEQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 QY 61 LITARDLGCALSHGGLLDLACAVEMVHAASLLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 DB 61 LITARDLGCALSHGGLLDLACAVEMVHAASLLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 QY 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSD 180
 DB 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSD 180
 QY 181 AILLTNQPKTSTLFCASTQMASIAANASCEARENLRHRSFLDLGQAFOLLDDLTGMDTGT 240
 DB 181 AILLTNQPKTSTLFCASTQMASIAANASCEARENLRHRSFLDLGQAFOLLDDLTGMDTGT 240
 QY 241 KDINQDACKSTLVNLLSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIOWFQKLA 300
 DB 241 KDINQDACKSTLVNLLSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIOWFQKLA 300
 QY 301 AVS 303
 DB 301 AVS 303

RESULT 3

AAO16020
 ID AAO16020 standard; protein; 303 AA.

XX AC AAO16020;

XX DT 20-FEB-2003 (first entry)

XX DE Pantoea stewartii geranylgeranyl pyrophosphate synthase.

XX KW Carotenoid; crt.

XX OS Pantoea stewartii.

XX PN WO200279395-A2.

XX PD 10-OCT-2002.

XX PF 25-JAN-2002; 2002WO-US002124.

XX PR 26-JAN-2001; 2001US-0264329P.

XX PR 04-MAY-2001; 2001US-0288984P.

XX PA (CRGI) CARGILL INC.

De Souza ML, Kollmann SR, May CA, Schroeder WA;

WPI; 2003-075455/07.
N-PSDB; APT14192.

Novel isolated nucleic acid useful e.g. to engineer host cells with the ability to produce particular carotenoids and polypeptides useful in cell-free systems to make particular carotenoids.

Claim 20; Page 61-62; 74pp; English.

The invention comprises the amino acid and coding sequence of a number of carotenoid (crt)-related proteins. The crt-related DNA and protein sequences of the invention are useful for engineering cells which are able to produce carotenoids. The present amino acid sequence represents a crt-related protein of the invention

Sequence 303 AA;

```
Query Match 99.5%; Score 1518; DB 6; Length 303;
Best Local Similarity 99.3%; Pred. No. 1.8e-151;
Matches 301; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 LTVCAKGVHLTGISAEQLADIDSRDLQPLVQGERDCVGAAREGTLPAGKRIKIRPMLL 60
1 MMTCAKGVHLTGISAEQLADIDSRDLQPLVQGERDCVGAAREGTLPAGKRIKIRPMLL 60
61 LITARDLGAISHGGLDLACAVENVHAASLILODMPDMDAQMREGPTIHTQYGEHVA 120
61 LITARDLGAISHGGLDLACAVENVHAASLILODMPDMDAQMREGPTIHTQYGEHVA 120
121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSA 180
121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSA 180
181 ALLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAAPOLLDDLTGMDTGT 240
181 ALLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAAPOLLDDLTGMDTGT 240
241 KDINDAGKSTLVNLGSGAVEERLRHPSLDLGAAPOLLDDLTGMDTGT 300
241 KDINDAGKSTLVNLGSGAVEERLRHPSLDLGAAPOLLDDLTGMDTGT 300
301 AVS 303
301 AVS 303
```

```
RESULT 4
AA07463
AA07463 standard; protein; 302 AA.
AA07463;
24-OCT-2003 (revised)
28-JAN-1991 (first entry)
Polypeptide with enzymatic activity for the conversion of prephytoene
Pyrophosphate into phytoene.
Carotenoid biosynthesis; vitamin A; cancer; food coloring.
Pantoea ananatis.
EP93690-A.
24-OCT-1990.
20-APR-1990; 90EP-00107493.
21-APR-1989; 89JP-00103078.
05-MAR-1990; 90JP-00053255.
```

(KIRI) KIRIN BEER KK.

Misawa N, Kobayashi K, Nakamura K;

WPI; 1990-322212/43.

N-PSDB; AAQ06293.

DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

Claim 1; Fig 1; 40pp; English.

Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

Sequence 302 AA;

```
Query Match 88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 3.9e-133;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;
1 LTVCAKGVHLTGISAEQLADIDSRDLQPLVQGERDCVGAAREGTLPAGKRIKIRPMLL 60
1 MMTCAKGVHLTGISAEQLADIDSRDLQPLVQGERDCVGAAREGTLPAGKRIKIRPMLL 60
61 LITARDLGAISHGGLDLACAVENVHAASLILODMPDMDAQMREGPTIHTQYGEHVA 120
61 LITARDLGAISHGGLDLACAVENVHAASLILODMPDMDAQMREGPTIHTQYGEHVA 120
121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSA 180
121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSA 180
181 ALLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAAPOLLDDLTGMDTGT 240
181 ALLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAAPOLLDDLTGMDTGT 240
241 KDINDAGKSTLVNLGSGAVEERLRHPSLDLGAAPOLLDDLTGMDTGT 300
241 KDINDAGKSTLVNLGSGAVEERLRHPSLDLGAAPOLLDDLTGMDTGT 300
301 AVS 303
301 AVS 302
```

RESULT 5
AA082255
AA082255 standard; protein; 302 AA.

AA082255;

17-OCT-2003 (revised)

16-JUL-1999 (first entry)

C. utilis crtE protein.

HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtE; carotenoid.

Pichia jadinii.

JP10248575-A.

22-SEP-1998.

12-MAR-1997; 97JP-00058012.

12-MAR-1997; 97JP-00058012.

(KIRI) KIRIN BREWERY KK.

R WPI; 1998-560727/48.
 R N-PSDB; AAV73179.
 X Gene useful for increase in carotenoid production - and preparation of
 T carotenoid.
 X

Example 2; Fig 5-7; 54pp; Japanese.

X This invention describes a novel method for the preparation of
 C carotenoids using genes and proteins isolated from *Candida utilis*. The
 C invention specifically describes the isolation of a 3-hydroxy-3-
 C methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence
 C represents the *Candida utilis* crtE protein which is used in the method of
 C the invention. (Updated on 17-OCT-2003 to standardise OS field)

X Sequence 302 AA;

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
 Best Local Similarity 88.4%; Pred. No. 3.9e-133;
 Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

```

Y 1 LTVCAKXVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAREGTLAPGKRIKIRPMLL 60
b :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MTVCAXKHVHLTDAAEQQLADIDRLDQLLPVEGERDVVGAAAREGALAPGKRIKIRPMLL 60
Y 61 LTTARDLGCAISHGGLLDLACAVEMVHAASLIIDDMPCMDAQMRRGRPTIHTQYGEHVA 120
b :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LTTARDLGCAVSHDGLLDLACAVEMVHAASLIIDDMPCMDAKLRGRPTIHSYGEHVA 120
Y 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPRSA 180
b :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ILAAVALLSKAFGVIAADAGLTPLAKNRAVSELSTAIGMGLVQGFQKLSGDKPRSAE 180
Y 181 AILLTNOFKTSTLFCASQMASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
b :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 AILLTNEFKTSTLFCASQMASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
Y 241 KDINQDAGKSTLVNLLGSGAVERLRQHLRLASEHLSAACQNGHSTTQLFIQAWFDKLLA 300
b :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 KDSNQDAGKSTLVNLLGSPRAVERLRQHLRLASEHLSAACQNGHGA-TQHPFIQAWFDKLLA 299
Y 301 AVS 303
b 300 AVS 302

```

RESULT 6

AAW87888 standard; protein; 302 AA.

AAW87888;

17-OCT-2003 (revised)

10-MAR-1999 (first entry)

Protein encoded by the carotenoid biosynthesis gene crtE.

Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;
 crtI gene; adonixanthin-3-glucoside; astaxanthin monoglucoside;
 carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
 food additive.

Pantoea ananatis.

JP10327865-A.

15-DEC-1998.

29-MAY-1997; 97JP-00140460.

29-MAY-1997; 97JP-00140460.

(KIRI) KIRIN BREWERY KK.

(KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

WPI; 1999-099030/09.

N-PSDB; AAV34080.

New carotenoid glucoside(s) - used as food additives.

Disclosure; Page 17-18; 26pp; Japanese.

The present sequence represents a protein involved in carotenoid
 biosynthesis. The specification describes astaxanthin diglucosides and
 adonixanthin-3'-glucosides. The specification also describes a method for
 the preparation of a carotenoid glycoside, in which all, or part of,
 carotenoid biosynthesis genes crtE, crtB, crtI, crtZ, crtX or crtW
 are introduced to a microbe or plant and expressed. The transformed
 organism is cultured and astaxanthin diglucosides, adonixanthin-3'-
 glucosides, and/or astaxanthin monoglucoside are collected. The
 carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003
 to standardise OS field)

Sequence 302 AA;

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
 Best Local Similarity 88.4%; Pred. No. 3.9e-133;
 Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

```

QY 1 LTVCAKXVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAREGTLAPGKRIKIRPMLL 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MTVCAXKHVHLTDAAEQQLADIDRLDQLLPVEGERDVVGAAAREGALAPGKRIKIRPMLL 60
QY 61 LTTARDLGCAISHGGLLDLACAVEMVHAASLIIDDMPCMDAQMRRGRPTIHTQYGEHVA 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LTTARDLGCAVSHDGLLDLACAVEMVHAASLIIDDMPCMDAKLRGRPTIHSYGEHVA 120
QY 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPRSA 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ILAAVALLSKAFGVIAADAGLTPLAKNRAVSELSTAIGMGLVQGFQKLSGDKPRSAE 180
QY 181 AILLTNOFKTSTLFCASQMASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 AILLTNEFKTSTLFCASQMASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
QY 241 KDINQDAGKSTLVNLLGSGAVERLRQHLRLASEHLSAACQNGHSTTQLFIQAWFDKLLA 300
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 KDSNQDAGKSTLVNLLGSPRAVERLRQHLRLASEHLSAACQNGHGA-TQHPFIQAWFDKLLA 299
QY 301 AVS 303
Db 300 AVS 302

```

RESULT 7

AAW99098 standard; protein; 302 AA.

AAW99098;

17-OCT-2003 (revised)

14-MAY-1999 (first entry)

Erwinia uredovora crtE protein sequence.

Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;

Pantoea ananatis.

JP11046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

257 GSGAVERLRQHLRLASHEHLSAACQNGHSTTQFLQAWFDKILA 300
 247 GADAAQKLRHIDSADKHLTFACPGGAIRO-FMHLWFGHLLA 289

RESULT 9
 AAW00171 standard; protein; 298 AA.

AAW00171;
 16-OCT-2003 (revised)
 25-MAR-2003 (revised)
 17-OCT-1996 (first entry)

N terminally modified geranylgeranyl pyrophosphate synthase.
 E. herbicola; geranylgeranyl pyrophosphate synthase; pARC376;
 GGPP synthase; biosynthesis; carotenoid; lycopene;
 farnesyl pyrophosphate; phytoene; fpp; isopentenyl pyrophosphate; IPP;
 tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;
 food colourant; herbicide; norflurazon.

Pantoea agglomerans.
 US5530189-A.
 25-JUN-1996.

22-JUL-1993; 93US-00096043.
 02-MAR-1990; 90US-00487613.
 18-MAY-1990; 90US-00525551.
 03-AUG-1990; 90US-00562674.
 28-FEB-1991; 91US-00662921.
 30-OCT-1991; 91US-00785568.

(STAD) AMOCO CORP.
 Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 Ausich RL;
 WPI; 1996-308824/31.
 N-PSDB; AAT37092.

DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for
 prodn. of lycopene, and to produce transgenic plants resistant to
 norflurazon.

Example 1; Fig 3; 87pp; English.

This sequence represents Erwinia herbicola geranylgeranyl pyrophosphate
 (GGPP) synthase encoded by DNA derived from the plasmid pARC489B. The
 enzyme encoded by this sequence differs from the wild type synthase in
 that 13 amino terminal amino acids are deleted and replaced by four
 extraneous amino acids. This causes the GGPP synthase to be more active.
 C-terminal truncation of the enzyme was found to further improve the
 activity. GGPP synthase is an enzyme which is involved in the
 biosynthesis of carotenoids, esp. lycopene, from the ubiquitous
 precursor, farnesyl pyrophosphate. In E. herbicola, phytoene has been
 found to be formed biosynthetically in a two-step process. The initial
 step is the condensation of farnesyl pyrophosphate (fpp) and isopentenyl
 pyrophosphate (IPP) to form GGPP. This reaction is catalysed by GGPP
 synthase. This first step is immediately followed by a tail to tail
 dimerisation of GGPP, catalysed by the enzyme phytoene synthase, to form
 phytoene. Lycopene is produced from phytoene by the catalytic action of
 phytoene dehydrogenase-4H. The genes encoding components of the lycopene
 biosynthesis pathway may be used to transform a host cell for the
 commercial production of lycopene which is used as a food colourant.
 Plants transformed with the phytoene dehydrogenase-4H coding sequence are
 protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to
 correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

SQ Sequence 298 AA;
 Query March 50.0%; Score 763.5; DB 2; Length 298;
 Best Local Similarity 56.0%; Pred. No. 9.4e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

QY 17 EQLLADIDSELDQLLPVQGERDCVGAAREGTLPAGKRIRPMLLLLTARDLGCALSHGGL 76
 DB 7 EVMRQSIDDHLAGLLPETDSQDIVSLAMREGVNAPKGRIRPLMLLAARDLRYQGSMP 66
 QY 77 LDIACAVEMFHAASLIIDDDMPCHDDAQMREGRTIHTOYGEHVAILAAVALLSKAFVIA 136
 DB 67 LDIACAVELHTASLMDDDMPCHDDAQMREGRTIHTOYGEHVAILAAVALLSKAFVIA 126
 QY 137 EAEGLTPIAKTRAVSELSTAIGMQLVQGVQGLSECDKPRSDAILLTNQFKSTILFCA 196
 DB 127 ATGDLPGERRAQAQVNELSTAVGLQGLVGGFRDLNDAALDRTPDAILSTNHLKTGILFSA 186
 QY 197 STOMASTAANASCARENHFRFSLDLQAOQLLDLTDGMDTDGKIDNODAGKSTLVNLL 256
 DB 187 MLQIVAIASASSPSTRETLHAFALDFQAFQLLDLDDDDHPETGKDRNKDAGKSTLVNRL 246
 QY 257 GSGAVERLRQHLRLASHEHLSAACQNGHSTTQFLQAWFDKILA 300
 DB 247 GADAAQKLRHIDSADKHLTFACPGGAIRO-FMHLWFGHLLA 289

RESULT 10
 AAW00342
 ID AAW00342 standard; protein; 298 AA.
 XX AC AAW00342;
 XX 16-OCT-2003 (revised)
 XX 25-MAR-2003 (revised)
 XX 10-DEC-1996 (first entry)
 XX DE N-terminally truncated geranylgeranyl pyrophosphate synthase.
 XX Wild type; N-terminally truncated; geranylgeranyl pyrophosphate; GGPP;
 XX GGPP synthase; phytoene synthase; carotenoid; phytoene; pARC308A;
 XX pARC489D; transit peptide; tobacco; chloroplast; skin disorder.
 XX Pantoea agglomerans.
 XX US5545816-A.
 XX 13-AUG-1996.
 XX 19-JUL-1993; 93US-00093577.
 XX 02-MAR-1990; 90US-00487613.
 XX 18-MAY-1990; 90US-00525551.
 XX 03-AUG-1990; 90US-00562674.
 XX 28-FEB-1991; 91US-00662921.
 XX 30-OCT-1991; 91US-00785569.
 XX (STAD) AMOCO CORP.
 XX Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 XX Ausich RL;
 XX WPI; 1996-383742/38.
 XX N-PSDB; AAT41742.
 XX Transformed plants containing DNA encoding Erwinia herbicola enzymes -
 XX esp. geranylgeranyl pyrophosphate synthase and phytoene synthase, allows
 XX large scale production of phytoene.
 XX Claim 1; Fig 3; 61pp; English.
 XX The sequences given in AAW00341-42 represent the wild type and an N-
 XX terminally truncated form of geranylgeranyl pyrophosphate (GGPP)

synthase. GGPP synthase is a enzyme which, along with phytoene synthase (see also AAW00343), catalyze the formation of carotenoids, particularly phytoene. The N-terminally truncated form of GGPP synthase has the first thirteen amino acids replaced by four heterologous amino acids derived from the plasmid PARC306A. This enzyme is about twice as active as the wild type enzyme. A C-terminal truncated protein was found to have even higher activity and was cloned into plasmid PARC489D. The GGPP synthase coding sequences may be attached to the transit peptide sequence given in AAW00344, for transport into tobacco chloroplasts. This allows the tobacco plants to produce high levels of phytoene for use in the treatment of skin disorders. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 298 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 298;
 Best Local Similarity 56.0%; Pred. No. 9.4e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 17 EQLLADIDRLDQLLPVQGERDCVGAAMEGTLAPGKIRPMLLLTARDLGCAISHGGL 76
 7 EVNRQSIDDHLAGLLPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMP 66
 77 LDLCACAVEMVHABSLIDDPMDQDQMRGRPTHTQYGEHVAILAVALLSKAFGVIA 136
 67 LDLCACAVELTHASLMDDPCMDNAELRGQPTTHKXFGESVAILASVGLLSKAFGLIA 126
 137 EABGLTPIAKTRAVSELSTAIGMQGLVQGFQDLSEGDKPRSDAAILTNQPKTSLFCA 196
 127 ATGDLPCERRAQAVNELSTAVGLQGLVGFQDLNDALDRTDAILSTNHLKTGLFSA 186
 197 STOMASIANASCEARENLRPSLDLQGFQFLLDITDCKTGTGKIDNODACKSTLVNLL 256
 187 MLQIVAIASASSPSTRETLHAFALDFQAFQGLLDLDLDDHPETGKDRNKDACKSTLVNRL 246
 257 GSGAVERLRQHLRLASEHLSAACQNGHSTTQFIQAFDCKKLA 300
 247 GADAAKQKLEHIDSADKELTTPACPOGGAIRO-FMHLWFGHILA 289

BULT 11
 #32470

AAW32470 standard; protein; 298 AA.

AAW32470;

17-OCT-2003 (revised)
 25-MAR-2003 (revised)
 15-JAN-1998 (first entry)

Erwinia herbicola geranylgeranyl pyrophosphate synthase.

Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;
 lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;
 yeast; plant; vitamin A; cancer.

Pantoea agglomerans.

Key Location/Qualifiers
 Misc-difference 91
 /note= "Encoded by ACC"

US5656472-A.

12-AUG-1997.

07-JUN-1995; 95US-00473512.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00526274.

28-FEB-1991; 91US-00662921.

21-JUL-1993; 93US-00095726.

XX (STAD) AMOCO CORP.
 .PA
 XX Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;
 PI Brinkhaus FL;
 XX
 DR WPI; 1997-414592/38.
 DR N-PSDB; AAT91542.
 XX
 PT DNA encoding Erwinia herbicola lycopene cyclase - for producing
 PT recombinant enzyme, and transgenic organisms with increased beta-carotene
 PT levels.
 XX
 PS Example 2; Fig 3; 102pp; English.
 XX

XX A novel DNA molecule has been isolated which encodes an Erwinia herbicola

XX lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA
 CC molecule comprises at least 1125 bp and is present in the plasmids
 CC PARC147, PARC1509, PARC1510 and PARC1520. The present sequence represents
 CC the amino acid sequence corresponding to the preferred heterologous (GGPP)
 CC structural gene of Erwinia herbicola geranylgeranyl pyrophosphate (GGPP)
 CC synthase. The new DNA molecule can be used to produce the recombinant
 CC enzyme and transgenic organisms, e.g. yeasts or plants, with increased
 CC beta-carotene levels. Beta-carotene is used as a colourant in margarine
 CC and butter and as an intermediate for vitamin A, and may prevent cancer.
 CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX

XX Sequence 298 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 298;
 Best Local Similarity 56.0%; Pred. No. 9.4e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

Qy 17 EQLLADIDRLDQLLPVQGERDCVGAAMEGTLAPGKIRPMLLLTARDLGCAISHGGL 76
 Db 7 EVNRQSIDDHLAGLLPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMP 66
 Qy 77 LDLCACAVEMVHABSLIDDPMDQDQMRGRPTHTQYGEHVAILAVALLSKAFGVIA 136
 Db 67 LDLCACAVELTHASLMDDPCMDNAELRGQPTTHKXFGESVAILASVGLLSKAFGLIA 126
 Qy 137 EABGLTPIAKTRAVSELSTAIGMQGLVQGFQDLSEGDKPRSDAAILTNQPKTSLFCA 196
 Db 127 ATGDLPCERRAQAVNELSTAVGLQGLVGFQDLNDALDRTDAILSTNHLKTGLFSA 186
 Qy 197 STOMASIANASCEARENLRPSLDLQGFQFLLDITDCKTGTGKIDNODACKSTLVNLL 256
 Db 187 MLQIVAIASASSPSTRETLHAFALDFQAFQGLLDLDLDDHPETGKDRNKDACKSTLVNRL 246
 Qy 257 GSGAVERLRQHLRLASEHLSAACQNGHSTTQFIQAFDCKKLA 300
 Db 247 GADAAKQKLEHIDSADKELTTPACPOGGAIRO-FMHLWFGHILA 289

RESULT 12

AAW01119
 ID AAW01119 standard; protein; 307 AA.

XX AAW01119;

XX 16-OCT-2003 (revised)
 XX 25-MAR-2003 (revised)
 DT 10-DEC-1996 (first entry)

XX Geranylgeranyl pyrophosphate synthase.

XX GGPP synthase, lycopene cyclase; phytoene synthase; beta-carotene;
 KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
 KW pigment; food colourant; chloroplast transit peptide; increase yield;
 KW tobacco ribulose bis-phosphate carboxylase-oxygenase.
 XX Pantoea agglomerans.

US5530188-A.
 25-JUN-1996.
 21-JUL-1993; 93US-00095726.
 02-MAR-1990; 90US-00487613.
 18-MAY-1990; 90US-00525551.
 03-AUG-1990; 90US-00562674.
 28-FEB-1991; 91US-00662921.
 30-OCT-1991; 91US-00785566.
 (STAD) AMOCO CORP.
 Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 Ausich RL;
 WPI; 1996-308823/31.
 N-PSDB; AAT40789.
 Increasing prodn. of total carotenoid(s) in a higher plant - by
 transforming with vector encoding chloroplast transit peptide operably
 linked to the Erwinia herbicola lycopene cyclase structural gene.
 Example 2; Fig 2; 99pp; English.
 The present sequence is that of the Erwinia herbicola geranylgeranyl
 pyrophosphate (GGPP) synthase. GGPP is a 20-carbon atom precursor of
 phytoene, the first carotenoid in the carotenoid biosynthesis pathway.
 The phytoene synthase (AAW01121) participates in the pathway by providing
 phytoene from GGPP. Production of total carotenoids in a plant can be
 increased by transforming the plant with DNA encoding enzymes involved in
 the biosynthesis pathway, in particular the lycopene cyclase gene
 (AAT40795). Lycopene is prepared biosynthetically from phytoene through
 four sequential dehydrogenation reactions which can be carried out by a
 single dehydrogenase (AAW01123) in Erwinia sp. . Beta-carotene is produced
 by the action of lycopene cyclase on lycopene. A chloroplast transit
 peptide of the tobacco ribulose bis-phosphate carboxylase- oxygenase gene
 (see AAT40794) is operatively linked in frame to the 5' end of the
 lycopene cyclase structural gene. This leads to increased production of
 total carotenoids in the chloroplast of transformed plants as compared to
 native, non-transformed plants of the same type. Beta- carotene is an
 effective and apparently harmless food colourant and is also in the
 pathway for biological synthesis of further C40 carotenoids such as
 zeaxanthin and zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct
 PP field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 307 AA;
 Query Match 50.0%; Score 763.5; DB 2; Length 307;
 Best Local Similarity 56.0%; Pred. No. 9.8e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 17 EQLLADISRLDQLPVQGERDCVGAAMREGLTAPGKIRPMLLLTARDLGCALSHGGL 76
 16 EWMRQSIDHLAGLPEPDSQIVSLWEGVWAGKIRPMLLLAARDLRYQSSMPTL 75
 77 LDLCACVEMVHAASLILDDMPQDAQWRGRPTHTQYGHVAILAAVALLSTAFGVIA 136
 76 LDLCACAVELTHTASLMDMPQCNDAELRGQPTTHKPGESVAILASVGLLSKAFGLIA 135
 137 EAGLTPIAKTRAVSLSTAIGMGIQVQGFNDLSEGGKPRSDAILTNQFTSTLPCA 196
 136 ATGDLPGERRAQAVALNELSTAVGLQGLVLQGFRLDNDALDTPDAILSTNHLKTLGILFSA 195
 197 STQMASIAANASCEAFENLHRESLDLQGFLLDLDTGWTDTGXDINQDAGKSTLVNLL 256
 196 MLQIVAIASASSPSTRETLHAFALDQGFLLDLDRDDHPETGKRNKDGKSTLVNRL 255
 257 GSGAVTEERLQHLRLASEHLSACQNGHSTTQLFIQAMFCKLA 300
 256 GDAARQKLEHIDSADKHLTACPGQGAIRQ-FMLHFWGHHLA 298

RESULT 13
 AAW00170
 ID AAW00170 standard; protein; 307 AA.
 XX
 AC AAW00170;
 XX
 DT 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 17-OCT-1996 (first entry)
 XX
 DE Geranylgeranyl pyrophosphate synthase.
 XX
 KW E. herbicola; geranylgeranyl pyrophosphate synthase; PARC376;
 KW GGPP synthase; biosynthesis; carotenoid; lycopene;
 KW farnesyl pyrophosphate; phytoene; FPP; isopentyl pyrophosphate; IPP;
 KW tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;
 KW food colourant; herbicide; norflurazon.
 XX
 OS Pantoea agglomerans.
 XX
 PN US5530189-A.
 XX
 PD 25-JUN-1996.
 XX
 PF 22-JUL-1993; 93US-00096043.
 XX
 PR 02-MAR-1990; 90US-00487613.
 PR 18-MAY-1990; 90US-00525551.
 PR 03-AUG-1990; 90US-00562674.
 PR 28-FEB-1991; 91US-00662921.
 PR 30-OCT-1991; 91US-00785566.
 XX
 PA (STAD) AMOCO CORP.
 XX
 PI Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 PI Ausich RL;
 XX
 DR WPI; 1996-308824/31.
 DR N-PSDB; AAT37091.
 XX
 XX DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for
 PT prodn. of lycopene, and to produce transgenic plants resistant to
 PT norflurazon.
 XX
 PS Example 2; Fig 2; 87pp; English.
 XX
 CC This sequence represents Erwinia herbicola geranylgeranyl pyrophosphate
 CC (GGPP) synthase encoded by DNA derived from the plasmid PARC376. GGPP
 CC synthase is an enzyme which is involved in the biosynthesis of
 CC carotenoids, esp. lycopene, from the ubiquitous precursor, farnesyl
 CC pyrophosphate. In E. herbicola, phytoene has been found to be formed
 CC biosynthetically in a two-step process. The initial step is the
 CC condensation of farnesyl pyrophosphate (FPP) and isopentyl pyrophosphate
 CC (IPP) to form GGPP. This reaction is catalysed by GGPP synthase. This
 CC first step is immediately followed by a tail to tail dimerisation of
 CC GGPP, catalysed by the enzyme phytoene synthase, to form phytoene.
 CC Lycopene is produced from phytoene by the catalytic action of phytoene
 CC dehydrogenase-4H. The genes encoding components of the lycopene
 CC biosynthesis pathway may be used to transform a host cell for the
 CC commercial production of lycopene which is used as a food colourant.
 CC Plants transformed with the phytoene dehydrogenase-4H coding sequence are
 CC protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to
 CC correct PP field.) (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 307 AA;
 Query Match 50.0%; Score 763.5; DB 2; Length 307;
 Best Local Similarity 56.0%; Pred. No. 9.8e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 17 EQLLADISRLDQLPVQGERDCVGAAMREGLTAPGKIRPMLLLTARDLGCALSHGGL 76

16 EVMRQSIDHLAGLIPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMTL 75
 77 LDLCACAVEMVHAASLILDDMPGDDAQRGRPTHTQYGEHVAILAAVALLSKAFGVIA 136
 76 LDLCACAVELTHTASLMLDDMPGDDAQRGRPTHTQYGEHVAILAAVALLSKAFGLIA 135
 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFKDLSEGDGKPSADAILTNQFKTSTLFCA 196
 136 ATGDLPGERRAQAQVNELSTAVGLQGLVLFQFRLDNDALDRTDPAILLSTNHLKTGILFSA 195
 197 STOMASIAANASCEARENHRSFSLDGOAFQFLLDLDLTDGMDTDTGKINDAGKSTLVNLL 256
 196 MLQIVAIASASSPSTRETTHAFALDFGQAFQFLLDLDLTDGMDTDTGKINDAGKSTLVNRL 255
 257 GSGAVEERLQHLRLASEHLSAACQNGHSTTQLFQAMFDDKLA 300
 256 GADARQKUREHIDSADKHLTFACPGGAIQO-FMELWFGHLLA 298

RESULT 14

AW00341

AAW00341 standard; protein; 307 AA.

AAW00341;

16-OCT-2003 (revised)

25-MAR-2003 (revised)

10-DEC-1996 (first entry)

Wild type geranyl pyrophosphate synthase.

Wild type; N-terminally truncated; geranylgeranyl pyrophosphate; GGPP;

GGPP synthase; phytoene synthase; carotenoid; phytoene; PARC306A;

PARC489D; transit peptide; tobacco; chloroplast; skin disorder.

Pantoea agglomerans.

US5545816-A.

13-AUG-1996.

19-JUL-1993; 93US-00093577.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

30-OCT-1991; 91US-00785569.

(STAD) AMOCO CORP.

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;

Ausich RL;

WPI; 1996-383742/38.

N-PSDB; AAT41741.

Transformed plants containing DNA encoding Erwinia herbicola enzymes -
 esp. geranyl pyrophosphate synthase and phytoene synthase, allows
 large scale production of phytoene.

Disclosure; Fig 2; 61pp; English.

The sequences given in AAW00341-42 represent the wild type and an N-

terminally truncated form of geranylgeranyl pyrophosphate (GGPP)

synthase. GGPP synthase is a enzyme which, along with phytoene synthase

(see also AAW00343), catalyze the formation of carotenoids, particularly

phytoene. The N-terminally truncated form of GGPP synthase has the first

thirteen amino acids replaced by four heterologous amino acids derived

from the plasmid PARC306A. This enzyme is about twice as active as the

wild type enzyme. A C-terminal truncated protein was found to have even

higher activity and was cloned into plasmid PARC489D. The GGPP synthase

CC coding sequences may be attached to the transit peptide sequence given in
 CC AAW00344, for transport into tobacco chloroplasts. This allows the
 CC tobacco plants to produce high levels of phytoene for use in the
 CC treatment of skin disorders. (Updated on 25-MAR-2003 to correct PF
 CC field.) (Updated on 16-OCT-2003 to standardise OS field)

XX SQ Sequence 307 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 307;

Sest Local Similarity 56.0%; Pred. No. 9.8e-72;

Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

QY 17 EQLLADISRLDQLPVGQERDCVCAANREGTLAPGKIRPMLLLTARDLGAISHGGL 76

Db 16 EVMRQSIDHLAGLIPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMTL 75

QY 77 LDLCACAVEMVHAASLILDDMPGDDAQRGRPTHTQYGEHVAILAAVALLSKAFGVIA 136

Db 76 LDLCACAVELTHTASLMLDDMPGDDAQRGRPTHTQYGEHVAILAAVALLSKAFGLIA 135

QY 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFKDLSEGDGKPSADAILTNQFKTSTLFCA 196

Db 136 ATGDLPGERRAQAQVNELSTAVGLQGLVLFQFRLDNDALDRTDPAILLSTNHLKTGILFSA 195

QY 197 STOMASIAANASCEARENHRSFSLDGOAFQFLLDLDLTDGMDTDTGKINDAGKSTLVNLL 256

Db 196 MLQIVAIASASSPSTRETTHAFALDFGQAFQFLLDLDLTDGMDTDTGKINDAGKSTLVNRL 255

QY 257 GSGAVEERLQHLRLASEHLSAACQNGHSTTQLFQAMFDDKLA 300

Db 256 GADARQKUREHIDSADKHLTFACPGGAIQO-FMELWFGHLLA 298

RESULT 15

AAW32469

ID AAW32469 standard; protein; 307 AA.

AAW32469;

25-MAR-2003 (revised)

15-JAN-1998 (first entry)

Geranylgeranyl pyrophosphate synthase.

Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;

lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;

yeast; plant; vitamin A; cancer.

Erwinia sp.

US5656472-A.

12-AUG-1997.

07-JUN-1995; 95US-00473512.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

21-JUL-1993; 93US-00095726.

(STAD) AMOCO CORP.

Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;

Brinkhaus FL;

WPI; 1997-414592/38.

N-PSDB; AAT91541.

DNA encoding Erwinia herbicola lycopene cyclase - for producing
 recombinant enzyme, and transgenic organisms with increased beta-carotene
 levels.

Example 2; Fig 2; 102bp; English.

A novel DNA molecule has been isolated which encodes an Erwinia herbicola lycopene cyclase enzyme that converts lycopene to beta- carotene. The DNA molecule comprises at least 1125 bp and is present in the plasmids PARC147, PARC1509, PARC1510 and PARC1520. The present sequence represents amino acid sequence corresponding to the nucleotide base sequences of certain preferred DNA segments of the structural gene for geranylgeranyl pyrophosphate (GGPP) synthase. The new DNA molecule can be used to produce the recombinant enzyme and transgenic organisms, e.g. yeasts or plants, with increased beta-carotene levels. Beta- carotene is used as a colouant in margarine and butter and as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-MAR-2003 to correct PF field.)

Sequence 307 AA;

Query Match	50.0%;	Score 763.5;	DB 2;	Length 307;
Best Local Similarity	56.0%;	Pred. No. 9.8e-72;		
Matches 159;	Conservative	40;	Mismatches 84;	Indels 1;
Gaps	1;			
17	ECLLADIDSLDQLLPVQGERDCVGAAMREGTLAPGKIRPMLLLLTARDLGCAISHGGL	76		
16	EVWQSIDDHLAGLLPETDSODIVSLAMREGVMAPGKIRPMLLLAARDLRYQGSMPFL	75		
77	LDLACAVEMVHAASLIIEDMPICMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFGVIA	136		
76	LDLACAVELTHTASLMLDDMPICMDNAELRGRQPTTHKKFGESVAILASVGLLSKAFGLIA	135		
137	BAEGLTPIAKTRAVSELSAIGMGLVQGFOLSEGDKPSADAILLTNQFTSTLPCA	196		
136	ATGGLPGERRAQAVNELSTAVGLQGLVIGQFRDLNDAALDRTPDALLSTNHLKTGLFSA	195		
197	STOMASTAANASCEARENLRFSLDLGOAFOLLDDLTGDMTDGKINODAGKSTLVNLL	256		
196	MLQIVATASASSPSTRTLHAPALDFQACQLDDLDHDPETGKDRNKDAGKSTLVNEL	255		
257	GSGAVEERLRQHLXSLASLHSAACONGHSTTQLFIQAWFDKXIA	300		
256	GADAAKQKLRHIDSADKHLTFACPQGGAIRQ-FMHLWFGHILIA	298		

Search completed: February 29, 2004, 14:43:51
Job time : 39.9246 secs

result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1337	51.5	494	16	Q9ENM2	Q9ENM2 oceanobacil
2	1287	49.5	497	16	Q9NR73	Q9NR73 staphylococ
3	1280	49.3	497	16	Q8NUQ3	Q8NUQ3 staphylococ
4	1050	40.4	439	2	Q53589	Q53589 staphylococ
5	888.5	34.2	534	16	Q7UJ66	Q7UJ66 rhodopirell
6	779	30.0	517	2	Q9ZGB4	Q9ZGB4 heliobacill
7	731	28.1	511	16	Q9ENM0	Q9ENM0 oceanobacil
8	724	27.9	492	2	Q8VUJ5	Q8VUJ5 pantosa agg
9	706	27.2	504	2	Q93CI7	Q93CI7 xanthobacte
10	687	26.4	492	2	Q8GCS1	Q8GCS1 pantosa ste
11	685	26.4	492	2	Q47845	Q47845 pantosa agg
12	672	25.9	502	16	Q9NR76	Q9NR76 staphylococ
13	672	25.9	502	16	Q8NUQ6	Q8NUQ6 staphylococ
14	667	25.7	506	2	O07855	O07855 staphylococ
15	662.5	25.5	512	17	Q9HPD8	Q9HPD8 halobacteri
16	658.5	25.3	506	2	O9KIX2	O9KIX2 bradyrhizob

Qy 127 QRF LDYSKN

127 QRF LDYSKNLCTET EAGYFAKGLDGFWDLLK FYGFLKRS
QY

100

127 ORFLDYSKNLCTETEAGYFAKGLDGFWDLLKFYGPLRSLLSFDVFRSMDQGVRRFISDPK 186

```

122 KSFLAYAKQIDDDITTKYDFDKGLDTTWELIIOHSGMLQSLKFKDLTSTWYEGIERITNPK 181
187 LVEILNFIKYVGSSPDAPALMNLIPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 246
182 LRNLAYFIKYVGSSPDAPALMNLIPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 241
247 RLDAEYSEIQKQD-GRACAVKLANGDVLPAIDVWSNMEVIPAMEKLRLSPASELKQM-QR 304
242 HTNSEVTKLAKDSTGNVIAATLADDSSEIKGDFISNMEVIPYKELMEKESYIKKLTKK 301
305 FPCSCGLVHLGVDRLYPOLAHNFFYSDHPRHFDVAFKSHRLSDPTIYLVAPCKTD 364
302 YEFSSGGLVHLGVDRLYPOLAHNFFYSDHPRHFDVAFKSHRLSDPTIYLVAPCKTD 361
365 PAQAPA-GCEIILKILPHIPLDPOKLLTAEDYSALRERVLVKLERMGLTDLRQHIIVTSEY 423
362 PNQVPGGVENIKILPHIPIQ-DKFFSDDDYKQFREQVLKLENKMGHGLRESIVTDM 420
424 WPLDIOAKYYSNQSIGYGVWADRFXNLGFKAPQSSLSNLYFYVGGSWNPGGKMPVTL 483
421 WTPNDIQSTYYSKSIYGTLSNCKINRGFKSPKSSKYDNLFFVGGSWNPGGKMPVTL 480
484 SGQVLRDKI 492
481 SGQVSEKI 489

```

RESULT 2

```

Q99R73 ID Q99R73 PRELIMINARY; PRT; 497 AA.
AC Q99R73;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein SAV2564.
GN SAV2564 OR SA2351.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RC SEQUENCE FROM N.A.
RP SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=2131952; PubMed=1148446;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayaashi H., Hiramatsu K.,
RA "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240 (2001).
DR EMBL; AP003365; BAB58726.1; -.
DR EMBL; AP003137; BAB43655.1; -.
DR PIR; E90061; E90061.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 497 AA; 57174 MW; E20BB9DDF5141C9D CRC64;

```

```

Query Match 49.5%; Score 1287; DB 16; Length 497;
Best Local Similarity 51.9%; Pred. No. 5.1e-93;
Matches 254; Conservative 94; Mismatches 143; Indels 2; Gaps 2;
QY 7 QRVTIVGAGLGSLAAISLATAGFSVQLIEKNDKVGKLNITWKDGTFFDLGSPILTMHP 66
DB 3 KHIIVGGGLGGISAAIRMAQSGYSVLSYEQNTHTGGKVRNHSDEGFGDLGSPILTMHP 62

```

```

QY 67 IFPALFTGAGNMADYVQIQKVEPHRNFFEDGSVIDLCEDAETQRELDKLGCGTYAQF 126
DB 63 IFPEKLFYSKQMSDYVITIKELPHQWLSFPDGTITDLVEGIKETQGHNAILSKQDIEEL 122
QY 127 QSLDYSKNLCTETAGYFARGDGFMDLLKFGPLRSLLSFDVFRSMQGVVRFISDEPK 186
DB 123 QWLYNYSRRIDRITEKGVNGLDLSQIIKFKHPLNALINYDVHTMQQAIDKRISNPY 182
QY 187 LVEILNFIKYVGSSPDAPALMNLIPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 246
DB 183 LRNLAYFIKYVGSSPDAPALMNLIPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 241
247 RLDAEYSEIQKQD-GRACAVKLANGDVLPAIDVWSNMEVIPAMEKLRLSPASELKQM-QR 304
242 HTNSEVTKLAKDSTGNVIAATLADDSSEIKGDFISNMEVIPYKELMEKESYIKKLTKK 301
305 FPCSCGLVHLGVDRLYPOLAHNFFYSDHPRHFDVAFKSHRLSDPTIYLVAPCKTD 364
302 YEFSSGGLVHLGVDRLYPOLAHNFFYSDHPRHFDVAFKSHRLSDPTIYLVAPCKTD 361
365 PAQAPA-GCEIILKILPHIPLDPOKLLTAEDYSALRERVLVKLERMGLTDLRQHIIVTSEY 423
362 PNQVPGGVENIKILPHIPIQ-DKFFSDDDYKQFREQVLKLENKMGHGLRESIVTDM 420
424 WPLDIOAKYYSNQSIGYGVWADRFXNLGFKAPQSSLSNLYFYVGGSWNPGGKMPVTL 483
421 WTPNDIQSTYYSKSIYGTLSNCKINRGFKSPKSSKYDNLFFVGGSWNPGGKMPVTL 480
484 SGQVLRDKI 492
481 SGQVSEKI 489

```

RESULT 3

```

Q8NUQ3 ID Q8NUQ3 PRELIMINARY; PRT; 497 AA.
AC Q8NUQ3;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE MW2485 protein.
GN MW2485
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamanoto K., Hiramatsu K.,
RA "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827 (2002).
DR EMBL; AP004830; BAB96350.1; -.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 497 AA; 57217 MW; 70814ED59449BBFF CRC64;

```

```

Query Match 49.3%; Score 1280; DB 16; Length 497;
Best Local Similarity 51.5%; Pred. No. 1.8e-92;
Matches 252; Conservative 86; Mismatches 149; Indels 2; Gaps 2;
QY 7 QRVTIVGAGLGSLAAISLATAGFSVQLIEKNDKVGKLNITWKDGTFFDLGSPILTMHP 66
DB 3 KHIIVGGGLGGISAAIRMAQSGYSVLSYEQNTHTGGKVRNHSDEGFGDLGSPILTMHP 62
QY 67 IFPALFTGAGNMADYVQIQKVEPHRNFFEDGSVIDLCEDAETQRELDKLGCGTYAQF 126

```



```
QY 346 SHRLSDPTIYLVAPCKTDPAPAGACBIIKILPHIHLDPDKLLTAEDYSAL-----RER 401
DB 378 KCQAPADPTAVCAPAISDPVAPDCESIYILVHTYLRP-----GHDWKAMLPGRDV 432
QY 402 VLVKLER-MGLTDLQHIIVTEYTFPLDIOAKYYSNQSIIYGVVADRFKNGKAPORSS 460
DB 433 ILDKLERTAGMBGLDAITVSDLSLTPEGIHNRVYVLANGAIYGLASHGKFTCAFPKGNRK 492
QY 461 ELSNLYFVGGSNPGGMPYTLSCQVLRDKIVADL 496
DB 493 DLHGLYLAGGAHPGPGMPVLMSCNIAANDSIDQDV 528

RESULT 6
Q9ZGE4 PRELIMINARY; PRT; 517 AA.
ID Q9ZGE4
AC Q9ZGE4
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Diapophytoene dehydrogenase CrtN.
GN CRTN
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
RL a major photosynthesis gene cluster from Helicobacillus mobilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR EMBL; AF080002; AAC84034.1; -.
DR PIR; T31463; T31463.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrxn_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phytin dehydro.
DR InterPro; IPR001100; Pyr redox.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00411; PNDRDTASEI.
DR ProDom; PDL39017; Phytin dehydro; 1.
SQ SEQUENCE 517 AA; 58340 MW; BB933977999C4587 CRC64;
```

```
Query Match 30.0%; Score 779; DB 2; Length 517;
Best Local Similarity 34.7%; Pred. No. 7.2e-53;
Matches 174; Conservative 107; Mismatches 196; Indels 24; Gaps 8;

QY 8 RVIVTGAGLGSLAALSATAGFSVOLIEKNDKVGKLNIMTKDGFDFLGPSTLTWPHI 67
DB 10 RVIVGSGAGGSAAVRLANQGVTVLEKATPGGRUSALQAGSITDVCGPTIMNDV 69

QY 68 FEALFTGAGKNADYVOIQKVEPHRNPFEDSGSIDLCEDAETORRELDKLGPGT----- 122
DB 70 FHOYFKDLGENTIDYLRVNPVPCVHLHTDGTCKPSIDLKELDLRSFNPDDVDGYL 129

QY 123 --YAFQRFQFLDYSKNLTETETAGYFAKGLDFWDLKLYGLRSLSPDFPSKXQGVRR 180
DB 130 RYLAQIHTHRYRYQAREKFIKES---FTKPSD-FFNITLIG----MMQLRTLNKYDDIAR 181

QY 181 FTSDPKLVILNYFYKYVGSSPYDAPALMNLIPYIOYH-GLWVYKGMYGMAQAMEKLA 239
DB 182 FIKDELRLSLTFQAILVGVSPFPDAPSYITLIGVYHGLSGVWPKGMAITQALVKLL 241

QY 240 VELGWVIRLDAVSEIQODGRACAVKLANGDVLPADIVVSNMEVTPAMEKLLRSPASEL 299
DB 242 GFQGSRLRVNAEVEQILIEQGSRAVGRVLANGVLEKADYVVISNADFPYTMENLV--PASHR 299
```

```
QY 300 -----KMQRFEPSCSLVHLGLVDRLYFQLAHNRFFVSDHPRHFDVAFKSHLSDDPT 354
DB 300 GKYTEQKLSNMEVTCGAPMLYGVNRRYDNLHLNIVFTPDYKYSMDLEFTRQQLPQDPA 359
QY 355 IYLVAPCKTDPAPAGACBIIKILPHIHLDPDKLLTAEDYSALRERVLVLERMGLTDL 414
DB 360 MYVREPTKYDOSVAPPGKDIILVLPVNL-SGIDWKETHRYRELVIKLERQGVTDL 418
QY 415 RGHIVTEBYTFLDITQAKYYSNQSIIYGVVADRFKNGKAPORSSLSNLYFVGGSNVP 474
DB 419 SKHIBFERIYTPETQNRFNFTYQGAAPGLAPSLFQSGVFRPHIKSKEVNLVFGASVHP 478
QY 475 GCGMPYTLSCQVLRDKIVAD 495
DB 479 GCGVPVVLVCGKLVSQVLAD 499

RESULT 7
Q8ENM0 PRELIMINARY; PRT; 511 AA.
ID Q8ENM0
AC Q8ENM0
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Phytoene dehydrogenase (phytoene desaturase) (EC 1.3.-).
GN OB2461
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RT Nucleic Acids Res. 30:3927-3935(2002).
RL EMBL; AP04601; BAC14417.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrxn_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR Oxidoreductase; Complete proteome.
SQ SEQUENCE 511 AA; 58463 MW; 8E70FD9F71D406AB CRC64;
```

```
Query Match 28.1%; Score 731; DB 16; Length 511;
Best Local Similarity 32.9%; Pred. No. 4.4e-49;
Matches 174; Conservative 122; Mismatches 161; Indels 72; Gaps 18;

QY 6 NORVIVGAGLGSLAALSATAGFSVOLIEKNDKVGKLNIMTKDGFDFLGPSTLTWPHI 65
DB 3 NKKVIVGAGVAGLASALRLCHAGENVVEYKESMPGSGHQITKQGTDFLGPSTVMP 62

QY 66 HIFEALFTGAGKNADYVOIQKVEPHRNPF-----ED-----GSVIDLCEDAETORRELD 116
DB 63 ELYREVVELTGRNPDYIPMERIDPMYNVFPFGNTPEDRYQISSDLIQLIDEIA-ISKD 121

QY 117 KLGPGTYAQ--PQRELDYSKNLCTETEAGYFAKGLDGF-----WDLKLYGP----- 161
DB 122 MAGFTYLNKLYQRF-----QIAK-DHFQIRFPHPW---QFTPTKLTWN 162

QY 162 ---LRSLSLFDVFRSMQGVRRFISDPKLVILNYFYKYVGSSPYDAPALMNLIPYIOYH 218
DB 163 GLKRLATLGNADNF-----INKYINDESLRQMSFQTLIYIGISPYNGPSLTIIPMSL 216

QY 219 YGLWVYKGMYGMAQAMEKLAVELGVEIRLDAVSEIQODGRACAVKLANGDVLPADIV 278
DB 217 YGVNVIKGMYTMATSMKRLFELGGTHYNSPVKVIIEKKATGKLENLEI-NADYV 275
```

```
279 VSNMEVDPAMEKLAFLSPASELK----KMQRPEPSCSGLVHLGVDRILYQLAH-HNFFYS 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
276 VCNADFPYSMNLIQDIKAKGKYTDKIDKMKYSCSCFIMVLGANKKYDEVSNVNFPS 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 DHPREHFAVFKSHRLSDDPYIYVAPCKTDPAPAGCEIILPHIPHLDPDKLLTA- 392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 ENLKETMDIDFTKRL-EDPAFYTVIGSKMDPSPSLAPEKGGIYVL--VPVSD---LSTAQ 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
393 -----EDYSALRERVVLKRMG-LTDLROKHIVTEYWTPLDIOAKYYSNQGSIYGVAD 446
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
390 YSWGDETAYTREKVFEXKLDEIDFKGIDKDEIVSETYWTPLDFESKNAYNGACFGRLPT 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
447 RFKNLGFAPQORSSLSNLYFVGGSVNPGGMMVTLGGQLVRDKIVAD 495
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
450 LQSNHLRPOSKAKCENLYFTGSSHPGAGVPIVLLSAXIATBELKLD 498

RESULT 8
V00J5
Q8VUJ5 PRELIMINARY; PRT; 492 AA.
Q8VUJ5;
01-MAR-2002 (TremBLrel. 20, Created)
01-MAR-2002 (TremBLrel. 20, Last sequence update)
01-JUN-2003 (TremBLrel. 24, Last annotation update)
CrtI protein.
Pantoea agglomerans pv. milletiae.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Pantoea.
NCBI_TaxID=182454;
[1]
SEQUENCE FROM N.A.
Kamunten H., Hirata R.;
"Isolation and characterization of carotenoid biosynthesis genes from
Pantoea agglomerans pv. milletiae wist 801."
Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AB076662; BAB79603.1; -.
GO; GO:0016491; P:oxidoreductase activity; IEA.
GO; GO:0003058; P:biocatalysis; IEA.
GO; GO:0008150; B:acetyl-CoA catabolic process; IEA.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
SEQUENCE 492 AA; 54802 MW; C9BE7DB0D936E31095 CRC64;
```

```
Query Match 27.9%; Score 724; DB 2; Length 492;
Best Local Similarity 33.5%; Pred. No. 1.5e-48;
Matches 166; Conservative 93; Mismatches 224; Indels 12; Gaps 5;

y 8 RVIVIGAGLGLSALISLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGSLTWP 67
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 3 RTTVIGAGFGGLALAIRLQASGVPTLLEQRDKPGGRAYVYQDQGTDFDAGT 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 68 RFALFTGAGKMWADYVQIKVPEWRFNPFEDGSVIDLCEDAETORRELDKLGPGY 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 63 IEELFTLAGKLSIDVELNPNVFPFRLCWESKVSFNDQPALEAQAFAFNPRDVEGR 122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 128 RFLDYSKNICTETETAGYFAKGLDGFWDLLKFGYPLRS---LLSPVFRSMQGVRR 184
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 123 RFLAYSGRAFAE---GYLKLGTVPF---LSFEDMLRAAPQLAKQAMRSVYSKVAS 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 185 PKLVELINLYFIKVGSSPYDAPALMNLPIYQHYGLWVYKGMVGMQAAMEKLAVEL 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 177 EHLRQAFSFLVLVGNPFPATSIYTLIHALERENGWVFPFGGTGALVQGMVKFLD 236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 245 EIRLDAEVESEIQDGRACAVKLANGDVLADIVVSNMEVIPAMEKLIRSPA---SE 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 237 EVELNASVARLEQENRITAVHLKDGVRFPFTRAVASNADVAVHTYRELLSQHPAS 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 302 MQRPEPSCSGLVHLGVDRILYQLAHNFFYSDHPRHFDAVFKSHRLSDDPYIYV 361
```

```
Db 297 LQNKEMSNLSFVIYVRELHNNHQAHTTVCFPYRELIDEIFNKDGLAEDFSYLHAPC 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 362 KTDPAQAPAGCEIILPHIPHLDPDKLLTAEDYSALRERVVLKERMGLTDLRQHI 421
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 VTDPSLAPBGGCSYVLAAPVRLGTADIDNAVEGPRLRDRIFDYLEQHYMPGLRS 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 422 EYWTPLDIOAKYYSNQGSIYGVADRFNKLGFAPQORSSLSNLYFVGGSVNPGG 481
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 RIPTTFDFRDELNAYQGSFSAFVEPILTQSAWFRPHNRDKNNILYLVGAGTH 476
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 482 TLSSQLVRDKIVADL 496
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 IGSAKATAGLMLEDL 491
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Q93CI7 PRELIMINARY; PRT; 504 AA.
Q93CI7;
01-DEC-2001 (TremBLrel. 19, Created)
01-DEC-2001 (TremBLrel. 19, Last sequence update)
01-JUN-2003 (TremBLrel. 24, Last annotation update)
Phytoene dehydrogenase.
CrtI.
Xanthobacter sp. (strain Pv2).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bacteri; Proteobacteria; Xanthobacter.
NCBI_TaxID=78245;
[1]
SEQUENCE FROM N.A.
Larsen R.A., Metcalf W.W.;
Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF408848; AAL02000.1; -.
GO; GO:0016491; P:oxidoreductase activity; IEA.
GO; GO:0003058; P:biocatalysis; IEA.
GO; GO:0006118; P:electron transport; IEA.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; B:acetyl-CoA catabolic process; IEA.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
SEQUENCE 504 AA; 55556 MW; CA848D11A607A828 CRC64;
```

```
Query Match 27.2%; Score 706; DB 2; Length 504;
Best Local Similarity 33.1%; Pred. No. 4.1e-47;
Matches 163; Conservative 87; Mismatches 237; Indels 6; Gaps 2;

Qy 7 QRIVIGAGLGLSALISLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGSLT 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 RRAAVIGSGFGGLSALIRLQAAGITTVFEQRDKPGGRAYVYQDQGTDFDGG 69
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 67 IFALFTGAGKMWADYVQIKVPEWRFNPFEDGSVIDLCEDAETORRELDKLGPG 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 CLEEVFAAAGKLSIDVELNPNVFPFRLCWESKVSFNDQPALEAQAFAFNPRD 129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127 QRELDYSKNICTETETAGYFAKGLDGFWDLLKFGYPLRSILSPVFRSMQGVRR 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 RFLAYSGRAFAE---GYLKLGTVPF---LSFEDMLRAAPQLAKQAMRSVYS 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 187 LVEILNLYFIKVGSSPYDAPALMNLPIYQHYGLWVYKGMVGMQAAMEKLAVEL 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 LRQAFSFLVLVGNPFPATSIYTLIHALERENGWVFPFGGTGALVQGMVKFL 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 247 RLDABVESEIQDGRACAVKLANGDVLADIVVSNMEVIPAMEKLIRSPA---SE 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 247 RLTSFVDEIWFQGRATAVKLKSGTILFPDIAVASNADVVHTYRELLRGAAR 306
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 304 REPPSCSGLVHLGVDRILYQLAHNFFYSDHPRHFDAVFKSHRLSDDPYIYVAP 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 KTRHSMNLSFVTFGARRTWDLQHTVLFPGPRYRGLVDEIFKGPNLDPDFSLY 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

364 DPAQAGCEIILKILPHILPHLDPKLLTADYDYSALRERVVLKLRMGLTDLRQHVITSEY 423
367 DKSLAPEGCTAFVVLSPFVHLGKADIDWEVGGPLYRDRILAHLEERLLAPGLDLSVTSRI 426
424 WPLDIOAKYNSNGSISYGVVADRFXNLGFKAPQSSLSNLYFYVGGSVNPGGEMWTL 483
427 LTFPGRFDELSAHCAGSAFSEPIILTSQAFRPHNRDKANLYFAGATHPGAGVGVVG 486
484 SGLVRDKIVADL 496
487 SAKATAGLILADL 499

RESULT 10
8GCS1 PRELIMINARY; PRT; 492 AA.
D Q8GCS1;
T 01-MAR-2003 (T-EMBLrel. 23, Created)
T 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Phytoene desaturase.
N CRTI.
S Pantoea stewartii.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=66269;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 8200;
A deSouza M.L., Kollmann S.R., Schroeder W.A.;
T "Carotenoid Biosynthesis (WO 02/079395 A2).";
L Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AY166713; AA855599.1; -.
R GO; GO:0016491; F:oxidoreductase activity; IEA.
R GO; GO:0009058; P:biosynthesis; IEA.
R GO; GO:0006118; P:electron transport; IEA.
R InterPro; IPR002937; Amino oxidase.
R Pfam; PF01593; Amino oxidase.
R PROSITE; PS00982; PHYTOENE DH; 1.
Q SEQUENCE 492 AA; 54836 MW; F3DCD224547ASFA CRC64;

Query Match 26.4%; Score 687; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.2e-45;
Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

364 DPAQAGCEIILKILPHILPHLDPKLLTADYDYSALRERVVLKLRMGLTDLRQHVITSEY 423
367 DKSLAPEGCTAFVVLSPFVHLGKADIDWEVGGPLYRDRILAHLEERLLAPGLDLSVTSRI 426
424 WPLDIOAKYNSNGSISYGVVADRFXNLGFKAPQSSLSNLYFYVGGSVNPGGEMWTL 483
427 LTFPGRFDELSAHCAGSAFSEPIILTSQAFRPHNRDKANLYFAGATHPGAGVGVVG 486
484 SGLVRDKIVADL 496
487 SAKATAGLILADL 499

RESULT 10
8GCS1 PRELIMINARY; PRT; 492 AA.
D Q8GCS1;
T 01-MAR-2003 (T-EMBLrel. 23, Created)
T 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Phytoene desaturase.
N CRTI.
S Pantoea stewartii.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=66269;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 8200;
A deSouza M.L., Kollmann S.R., Schroeder W.A.;
T "Carotenoid Biosynthesis (WO 02/079395 A2).";
L Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AY166713; AA855599.1; -.
R GO; GO:0016491; F:oxidoreductase activity; IEA.
R GO; GO:0009058; P:biosynthesis; IEA.
R GO; GO:0006118; P:electron transport; IEA.
R InterPro; IPR002937; Amino oxidase.
R Pfam; PF01593; Amino oxidase.
R PROSITE; PS00982; PHYTOENE DH; 1.
Q SEQUENCE 492 AA; 54836 MW; F3DCD224547ASFA CRC64;

Query Match 26.4%; Score 687; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.2e-45;
Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

360 PSLAPEGCSYVYLAVPVHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVTHRMF 419
425 TPLDIOAKYNSNGSISYGVVADRFXNLGFKAPQSSLSNLYFYVGGSVNPGGEMWTL 484
420 TPFDFDELNAWQSAFSEPIILTSQAFRPHNRDKANLYFAGATHPGAGVGVVG 479
485 GOLVRDKIVADL 496
480 AKATAGLILADL 491

RESULT 11
Q47845 PRELIMINARY; PRT; 492 AA.
AC Q47845;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93371414; PubMed=8395826;
RA Liu S.T.;
RT "Carotenoid-biosynthesis genes as a genetic marker for the purpose of
gene cloning.";
RL Biochem. Biophys. Res. Commun. 195:259-263(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94236237; PubMed=8180698;
RA To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
Chang Y.S., Liu S.T.;
RT "Analysis of the gene cluster encoding carotenoid biosynthesis in
Erwinia herbicola Eho13.";
RL Microbiology 140:331-339(1994).
DR EMBL; M90698; AAA21263.1; -.
DR PIR; S52586; S52586.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac_phytoene_dh.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;

Query Match 26.4%; Score 685; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.8e-45;
Matches 164; Conservative 90; Mismatches 226; Indels 12; Gaps 5;

11 VIGAGLGLSAAISLATAGSFVOLIEKNDKVGKLNMTKDGTFDLGSLTMTMHPHFEA 70
6 VIGAGLGLALALRLQAAGIPVLLQRQKPGRAVYEDQGFEDAGPTVITDPSAIE 65
71 LFTGAGKGMADYVQIQKVEPHWRNFFDGSVIDLCDAETQRRELKLGPGTYAQFREL 130
66 LFTLAGQLADYVELLPVTPFYLWCWESKVFVNDQALQAEQIQQFNPDRVGEYRQL 125
131 DYSKNLCTETAGYFAKGLDGFWDLLKFGYPLRS---LLSPDFVSRMDQGVRRFISDPKL 187
126 DYSRAVFE---GVLKLGTVFP---LSFRDLRAAPQLAKLQAWRTVYSKVASYIEDEHL 179
188 VEILNYFIKYVGGSPYDAPALMNLPIYQHYGLWYVKGMYGMAQMEKLAIVELGVEIR 247
180 RQAFPSHLLVGGNPFATSIYTLIHALEREWGVPFRGGTGAIVNGMKILFQDLGGEVV 239
248 LDAEVSIIQDQGRACAVKLANGDVLPAIVSNMEVIPAMEKLL-RSPAS--ELAKQQR 304

```

1 240 LNAYSHMETTGTDTIEAHELEDGRFFPRVASNADVVHTYDRLLSQHPAAVKOSKKLOT 299
2 305 FEPSCGVLVHLGVDRLYPOLAHENFFVSDHPRHPFPAVKSHRLSDDPYIYLVAPCKTD 364
3 300 KRMSNLFVLYFLGHHHDQLAHHTVCFGRPYRELIHFIHNDGLADDFSLYLHPACVTD 359
4 365 PAQAPAGEI1KILPHIPELOPKLLTAEDYSALRERVLVXLERMGLTDLRQHIVTVEYW 424
5 360 SSLAPEGCGSVYVAPVPHLG.TANLDWTEGPRLRDRAIFEVLEGHYMPGLRSQVLTQRMF 419
6 425 TPLDIOAKYNSQSGIYGVADRFKNLOFKAPORSSLSNLIYFVGGSVNPGGMPWNTLS 484
7 420 TPFDPDQINAYQGSFAVEPVLTQSAMFRPHNRDKTNNLYLVAGTHPGAGTGPVIGS 479
8 485 GOLVZDKIVADL 496
9 480 AKATAGLMEDL 491

```

```

1 9876
2 9876 PRELIMINARY; PRT; 502 AA.
3 01-JUN-2001 (TREMELrel. 17, Created)
4 01-JUN-2001 (TREMELrel. 17, Last sequence update)
5 01-OCT-2003 (TREMELrel. 25, Last annotation update)
6 Squalene synthase.
7 CRTN OR SAV2561 OR SA2348.
8 Staphylococcus aureus (strain Mu50 / ATCC 700699), and
9 Staphylococcus aureus (strain N315).
10 Bacteria; Firmicutes; Bacillales; Staphylococcus.
11 [1]
12 NCBI_TaxID=158878, 158879;
13 SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
14 MEDLINE=21311952; PubMed=11418146;
15 Kuroda M., Ouchi A., Aoki K.-I., Negai Y., Lian J.-Q., Ito T.,
16 Cui L., Oguchi A., Aoki K.-I., Negai Y., Lian J.-Q., Ito T.,
17 Kanamori M., Matsumaru H., Maruyama A., Murakami H., Haseyama A.,
18 Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
19 Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
20 Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
21 Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
22 "Whole genome sequencing of methicillin-resistant Staphylococcus
23 aureus.";
24 Lancet 357:1225-1240(2001).
25 EMBL; AP003365; BAB58723.1; -.
26 EMBL; AP003337; BAB43652.1; -.
27 PIR; B90061; B90061.
28 GO; GO:0006118; P:electron transport; IEA.
29 InterPro; IPR000759; Adrnx_reductase.
30 InterPro; IPR002937; Amino_oxidase.
31 InterPro; IPR002055; NAD_BS.
32 InterPro; IPR008151; Phyt_n_dehydro.
33 Pfam; PF01593; Amino_oxidase; 1.
34 PRINTS; PR00419; ADXEDTASE.
35 ProDom; PD139017; Phyt_n_dehydro; 1.
36 Complete proteome.
37 SEQUENCE 502 AA; 56742 MW; 969528497B394203 CRC64;

```

```

1 Query Match 25.9%; Score 672; DB 16; Length 502;
2 Best Local Similarity 31.5%; Pred. No. 2e-44;
3 Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;
4 8 RVIVTGAGLGGLSAISLATAAGFSVQLTEKNDKVGKLNIMTKDGFEDLGPSTLTPHI 67
5 2 KIATVAGTGTGLAAARASQGHETIIFKNNVGGRNQKKGFTFDMGFTIVMPDV 61
6 68 FEALFTGAGKNADYVQIQVOKVPEHNRNFFDGSVIDLCEDAETORRELDKLGPGTYAQFQ 127
7 62 YKDVFTACGNVEDYIELRQLRYIVDVYFDHDDRTITVPTDLAEQLQOMLESIEPGSTHGM 121

```

```

1 128 RPLD--YSKNLCTETEAGYF-----AKGLDGFMDLLKFGFLRSLLSFDVFRSM---DQG 177
2 122 SFLTDVYK---YEIARRYFLERTYRKPSD-----FYN-MTSLVQAKLKTILNHADQL 170
3 178 VRRFTSDPKLVEILNFFIKYVSSPYDAPALMNLPLPYIOHYGLWYVKGWYGAQAMEK 237
4 171 IEHYDNEKIOKLAFOFLYIGIDPKRGPSLYSIIPMIEMFVGHFKGMYGHAQGLAQ 230
5 238 LAVELGVIRLDAEYSE--IQQDGRACAVKANGDVLPAIVVSNMEVIPAMEKLLRSP 295
6 231 LNKDLGVNIELNAEIEQIIIDPKFRADAIKV-NGDIRKFDKILCTADFPSSVAESLMPDF 289
7 296 ASEELK-----KMQREPEPSCGLVHLHGVLD-RLYPOLAHNNPFYSDHPRHPFPAVKSHRLS 350
8 290 APIKYPPEKIAADLDYSCAFMLYIGIDIDVTDQVRLHNVIISDDFRGNIEBEIPEG-RLS 348
9 351 DDPTIYLVAPCKTDPAPAGCEI1KILPHIPLHD-----PDKLLTAEDYSALRERVL 403
10 349 YDPSIYVYVPAVDKSLAPEGKTYIYVLMPTPELKTSGSIDWSDEALTQQ-----IKEIY 404
11 404 VKLERMGL-TDLRQHIVTEYWTPLDIOAKYNSQSGIYGVADRFKNLGFKAQORSEL 462
12 405 RKLATIEVFEDIKSHIVSETIFTNDFEQTYHAKFGSAFGLMPTLAQSNYYRPNQVSRDY 464
13 463 SNLIYFVGSGVNPGGMPWNTLSGQLVRDKIVADIQ 497
14 465 KDLYFAGASTHPGAGVPIVLSAKITVDENIKDIE 499

```

RESULT 13

```

1 Q8NUO6 PRELIMINARY; PRT; 502 AA.
2 Q8NUO6;
3 DT 01-OCT-2002 (TREMELrel. 22, Created)
4 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
5 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
6 DE Squalene synthase.
7 GN CRTN OR MW2482.
8 OS Staphylococcus aureus (strain MW2).
9 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
10 OC NCBI_TaxID=158620;
11 RN [1]
12 SEQUENCE FROM N.A.
13 RX MEDLINE=22040717; PubMed=12044378;
14 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
15 Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
16 Yamamoto K., Hiramatsu K.;
17 RT "Genome and virulence determinants of high virulence community-
18 acquired MRSA.";
19 RL Lancet 359:1819-1827(2002).
20 DR BMBL; AP004830; BAB96347.1; -.
21 DR GO; GO:0006118; P:electron transport; IEA.
22 DR InterPro; IPR000759; Adrnx_reductase.
23 DR InterPro; IPR002937; Amino_oxidase.
24 DR InterPro; IPR002055; NAD_BS.
25 DR InterPro; IPR008151; Phyt_n_dehydro.
26 DR Pfam; PF01593; Amino_oxidase; 1.
27 DR PRINTS; PR00419; ADXEDTASE.
28 DR ProDom; PD139017; Phyt_n_dehydro; 1.
29 KN Complete proteome.
30 SEQUENCE 502 AA; 56714 MW; 969529997B394203 CRC64;

```

```

1 Query Match 25.9%; Score 672; DB 16; Length 502;
2 Best Local Similarity 31.5%; Pred. No. 2e-44;
3 Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;
4 8 RVIVTGAGLGGLSAISLATAAGFSVQLTEKNDKVGKLNIMTKDGFEDLGPSTLTPHI 67
5 2 KIATVAGTGTGLAAARASQGHETIIFKNNVGGRNQKKGFTFDMGFTIVMPDV 61
6 68 FEALFTGAGKNADYVQIQVOKVPEHNRNFFDGSVIDLCEDAETORRELDKLGPGTYAQFQ 127
7 62 YKDVFTACGNVEDYIELRQLRYIVDVYFDHDDRTITVPTDLAEQLQOMLESIEPGSTHGM 121

```



```
{ PRINTS; PRO0368; PBDPNE.
{ ProDom; PDI39017; Phytin dehydro; 1.
{ PAD; Flavoprotein; Oxidoreductase; Complete proteome.
} SEQUENCE 512 AA; 56274 MW; 95C75E5A05F9126 CRC64;

Query Match      25.5%; Score 662.5; DB 17; Length 512;
Best Local Similarity 32.4%; Pred. No. 1.1e-43;
Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;

9 VIVIGAGLGGLSAASISLATAGFSVOLIEKNKVGKLNIMTKDGTDFDLGPSILTMPIHF 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9 VAVIGAGFGGLSTACYLADAGADVTVVEKTDIGGRASLTLEKDFRFDGMSWYLMEDVF 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

69 EALFTGAGKNMADYVQICKVSPHNEFE-----DGSVIDLCEDAET 110
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 EEPFASFDSPSYGLTRDPHRAIFKONEGRPGEDAPGLNVDTGGTIDVTPDREQ 128
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

111 QRRELDKLGPGTYAQFRLPYSKNECTETBAGYFAKGLDGF-----WDLKPYG 160
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 VKQVFDAYEPGAGDVLDDYLAQAKE-----NVEVGMEHFVKTDRPRVDMWDPKLA 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

161 PLRSLLSFDVFRSMDOGYRRFISDPKLVIELNYFIKYVGSSPYDAPALMNLIPVIOHYG 220
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 YARGL---TLGSMQDHVEQYDFRFLQVNVQYTLVFLGGSFDTTPALYNLASHVDFGLG 237
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

221 LMVYKGMVYGAQAAMEKLAVELGVEIRLCAEVSETQODGRACAVKLAN--GDVLPADIV 278
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 VYYPGGMGVVVDGITALARELGVEFRGHPVSAIK--GRRGGFKLDTBPGDAVLADV 294
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

279 VSNMEVITAMEKLL---RSPASELKQWRFEPSCSGLVHLGVDRLYPOLAHENFFYSD 334
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 VSDADYAHTEQALLPAKQKQYDADYWDARTYAP--SAFLLYLGVGDEVEPLAHTLVLP 352
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

335 HPREHFDVAFKSHRLSDPTIYLVAPCKTDPAPAG-CEIKILPHIPHLDPDKLLTAE 393
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 DWDTHFAQIFDDPAPNPDPAAYLCVPSKTDDTVAPDGHSLFALVPVAPGLD-DTPAVRE 411
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

394 DYSALRERVVLKRLKELTDLRQHLVTEBYWTPLDIOAKYYSNQGSIYGVVADRFKNLGF 453
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 SY---RDLVLDLIAENTGVLDLDRIVVEERFSVSEFADRYNSHQGTALGLAHTLRQTALF 468
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

454 KAPQSSSELNLYFVGGSVNPGGMPVLTLSGLVRDKIVAD 495
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 RPPHRSALDGLYFTGANTTPOIGVPMCLISGEVTANNVLAD 510
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: February 25, 2004, 14:50:45
Job time: 42.8667 secs

Result No.	Query			ID	Description
	Score	Match	Length		
1	773	29.8	517	CRTU MYXXA	P54979 myxococcus
2	770	29.6	492	CRTI ERWHE	P22871 erwinia her
3	671	25.8	492	CRTI PANAN	P21685 pantoea ana
4	645	24.8	501	CRTI AGRAU	P54978 agrobacteri
5	641.5	24.7	583	CRTI_PHYEL	P54982 phycocyc
6	638	24.6	529	CRTI MYXAX	Q02861 myxococcus
7	631.5	24.3	621	CRTI CERNC	P48537 cercospora
8	624.5	24.0	595	CRTI NEUCR	P21334 neurospora
9	581.5	22.4	524	CRTI RHUCA	P17054 rhodobacter
10	548	21.1	518	CRTI RHOSH	P54980 rhodobacter
11	541.5	20.8	495	CRTD RHOSH	Q01671 rhodobacter
12	526.5	20.3	507	CRTI_STRGR	P54981 streptomyc
13	514.5	19.8	508	CRTI_STRSE	P54971 streptomyc
14	490.5	18.9	494	CRTD RHUCA	P17059 rhodobacter
15	174.5	6.7	472	CRTI_SYNY3	P29273 synecocyst
16	158.5	6.1	570	CRTI_NARPS	Q40406 narcissus p
17	157	6.0	516	OXLX_CROAD	O93364 crotalus ad
18	156	6.0	519	AOFB_RAT	P19643 rattus norv
19	152.5	5.9	570	CRTI_SOYBN	P29553 glycine max
20	150	5.8	500	PAO_MAZE2	O64411 zea mays (m
21	149.5	5.8	566	CRTI_ARATH	Q07356 arabidopsis
22	148.5	5.7	474	CRTI_SYNP7	P26294 synecococc
23	146	5.6	469	P49_STRLI	P06108 streptomyc
24	145.5	5.6	535	P397_MYCTU	Q10555 mycobacteri
25	144.5	5.6	478	PUC_MICRU	P40974 micrococcu
26	144	5.5	500	PCLI_ARATH	P57681 arabidopsis
27	143.5	5.5	571	CRTI_MAZE1	P49086 zea mays (m
28	142.5	5.5	566	CRTI_ORYSA	Q92tn9 oryza sativ
29	142	5.5	508	FMS1_YEAST	Q02664 saccharomyc
30	139.5	5.4	583	CRTI_LYCES	P28554 lycopersico
31	136.5	5.3	522	AOFP_ONCMY	P49253 oncorhynch
32	135	5.2	544	A37C_DROLE	Q06570 drosophila
33	134	5.2	630	FIG1_MOUSE	Q09046 mus musculu

DR 4 STQGRIVVVGAGVGLAAARLAHQGFQVVEKTOGPOGRNRLQVQGFMDLQPTIV 63
DR 63 TMRPHIPALFTGAGKXNMDVYQIKVZPHWRNPFDSGVIDLCEAETORRELDKLGRC 122
DR 64 LMPEVPEFTFVAVGRRIEDVLTLLRCDFNRYVHFRDRSDVTFTELCAHRELERVEPS 123
DR 123 YAOQFORPLDYSKNLCTETETAGYFAKGLDGFWDLLKFGPLRLSLLS-----FDV--PFS 173
DR 124 YARVLAFLAQR-----VQVRTSLDEL--VGRNVAGLRDYLSPVLARIFOVRAHR 173
DR 174 MDQGVRRPISDPKLVILNFIKTVKSPDAPALMNLPIYQHYHGLVAVKGMVMAQ 233
DR 174 MYADVSREFFODERLAAMTFTQVLYGSPASPAVYGLLPFTLGLGVIMFPKGLVAIFQ 233
DR 234 AMEKLAVALGVEIRLDAEVSIOQDGRACAVKLANGDVLPAQIVVSNMEVIVPAMEKLR 293
DR 234 ALERLARESGVRYHGAVERILTDGRTGVRLEGSEVVEADAVLCNADLPAYEKLID 293
DR 294 SPASELKQORPEPSCGLVHLGVDRLLPOLAHNFFYSDHREHDAVFKSHRLSDDP 353
DR 294 PRATTLKREKRLVTSSTGYMLYGMKARYPELLHHNVVFCRDYKGSFDDIFE-FRVPEDP 352
DR 354 TIYLVAPCKTDPAPAGACEIIKILPHILDPDKLLTAEDYSALRERVLVLERMGLTD 413
DR 353 SFYVNAPTRTDASLAPEGKDALYVLPVPHQHPDLKWKVEG-PKVRAPKFAVAAELGFPS 411
DR 414 LRQHI-VTEEXWTPLDIOAKYXSNQSGIYGVVADRFKNLG-FKAPQSSLSNLYFVGG 471
DR 412 LESDIEVRRSSTPDDMAGTFLNARGSGFG-LSQNFQIQPFPSNQDARVKNLFFVGAS 470
DR 472 VNPGGMPVTLSSGQLVRDKIV 493
DR 471 TOPGGTLPVLISARLVTERLX 492

RESULT 2

DR CRTI_ERWHE STANDARD; PRT; 492 AA.
AC P22871;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SHO10;
RX MEDLINE=9108634; PubMed=2263648;
RA Armstrong G.A., Alberti M., Hearst J.E.;
RT "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes."; Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediaries of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds.
CC -!- COFACTOR: PAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; M38423; AAA24820.1; -

DR EMBL; M87280; AAA64981.1; -.
DR PIR; A39273; A33120.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt. dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR Prodom; PD139017; Phyt. dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; PAD; Flavoprotein; NAD.
FT NP_BIND 5 38 PAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8BDCSDB1562083F2 CRC64;
Query Match 29.6%; Score 770; DB 1; Length 492;
Best Local Similarity 34.7%; Pred. No. 3.5e-51;
Matches 174; Conservative 93; Mismatches 213; Indels 22; Gaps 6;
QY 7 QRIVVIGAGLGGLSAAALSLATAGSVOLIEKNKVGKLNIMTKGTFOLGHSILTMPH 66
DB 2 KKTVVIGAGFGGLALAIRLQAAGIPTVLEQRDKGGRAVYVHDDQGGTFQAGPTVITDPT 61
QY 67 IFEALFTGAGKXNMDVYQIKVZPHWRNPFDSGVIDLCEAETORRELDKLGFGTYAQF 126
DB 62 ALEALFTLAGRMEDYVRLLPVKFYLCHESGKTLDYANDSAELEAQITQFNPRDVEGY 121
QY 127 QRFLDYSKNLCTETETAGYFAKXG---LDGFMDLLKFGYPLRLSLLSFDVFRSMDQGVRRFIS 183
DB 122 RFLAYSQAVFOE---GYLRGVSVPFLSFRDMLR-AGP--QLLKLAQWSYQSVSRFIE 175
QY 184 DPKLVILNLYIKVGVSPYDAPALMNLPIYQHYHGLVAVKGMVMAQAMEKLAVALG 243
DB 176 DEHLRQAEPFSLHLLVGGNPFPTSSIIYLIHALERENGWVPEGGTGALVNGVXLFTDLG 235
QY 244 VEIRLDAEVSIOQDGRACAVKLANGDVLPAQIVVSNMEVIVPAMEKLL-----RSP 295
DB 236 GSIELNARVEELVADNRVSOVRLADGRI FDTDAVASNADVNTYKLLGHHHPYGOKRAA 295
QY 296 ASELKQORPEPSCGLVHLGVDRLLPOLAHNFFYSDHREHDAVFKSHRLSDDP 355
DB 296 ALERKSM-----SNSLFVLYFGLNPHSOLAHHTTCFGRYRELIDEIFTGSAADDFSL 350
QY 356 YLVAPCKTDPAPAGACEIIKILPHILDPDKLLTAEDYSALRERVLVLERMGLTDLR 415
DB 351 YLHSCVYTDPSLAPPGCAFVFLAPVPHLGNALPDWAQEGPKLRDRIFDYLEERYMPGLR 410
QY 416 QHIVTEEXWTPLDIOAKYXSNQSGIYGVVADRFKNLGFKAPQSSLSNLYFVGGSNPG 475
DB 411 SOLVQRTFTPADFHDTLDAHLGSAFSTIEPLLTQSAWFRPHNRDSITANLYLVGATHPG 470
QY 476 GGPMPVTLSSGQLVRDKIVADIQ 497
DB 471 AGIPGVVASAKATASLMIEDLQ 492

RESULT 3

DR CRTI_PANAN STANDARD; PRT; 492 AA.
AC 221685;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
RA Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y., Nakamura K., Harashina K.;
RA "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway

by functional analysis of gene products expressed in *Escherichia coli*."

J. Bacteriol. 172:6704-6712(1990).

-!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediaries of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds.

-!- COFACTOR: FAD (Probable).

-!- PATHWAY: Carotenoid biosynthesis.

-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D90087; BAA14127.1; -

PIR; D37802; D37802.

InterPro; IPR002937; Amino oxidase.

InterPro; IPR008150; Bac_phytoene_dh.

InterPro; IPR008151; Phyt_n_dehydro.

Pfam; PF01593; Amino oxidase; 1.

ProDom; PD139017; Phyt_n_dehydro; 1.

PROSITE; PS00982; PHYTOENE_DH; 1.

Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.

NP_BIND 5 38 FAD (ADP PART) (POTENTIAL).

SEQUENCE 492 AA; 55007 MW; FAACD4E34A9C6413 CRC64;

Query Match 25.8%; Score 671; DB 1; Length 492;

Best Local Similarity 32.7%; Pred. No. 1.3e-43;

Matches 161; Conservative 89; Mismatches 230; Indels 12; Gaps 5;

11 VTGAGLGLSAAISLATAGFSVQLIKNDKVGKLNIMTKDGTFDLGPSTLTMPHIFE 70

6 VTGAGFGGLAALRLQAGIPVLLQEKDKPGGRAYVEDGTFDAGTPTVITDPSAIE 65

71 LFTGAGKNMADYVQIQKVEPHWRNFFEDGSVIDLCEDAETORRELDKLGPGTYAQFOR 130

66 LFLAGKQLKEVVELLPVTPPYRLCWESGKVFVNDQTRLEAQIQQFNPDRDVBGRQFL 125

131 DYSKNLCTETAGYPAKGLDGFNDLLKFYGLPS---LISFDFVRSMDQGVRRFISDPKL 187

126 DYSRAVFKD---GYLKLGTVPF---LSFRDMLRAAPQLAKQAWRSVTSKVASIRBEHL 179

188 VEILNFIKYVGSSPYDAPALMNLPPYIQHYGLVYVKGMYGMAQAMEKLAVELGVEIR 247

180 ROAFSPHSLVGNFPATSSIVTLIHALERENGWVFRGGTGALVQGMKILFQDLGGEW 239

248 LQAEVSEIQKQGRACAVKLANGDVLDPADIVVSNMEVTPAMEKLL-RSPAS--ELKKMQR 304

240 LNAVYSHMETTGNKIEAVHELDGRFLTAQVASNADVHTYREDLLSQHPAAVKOSNKLOT 299

305 FEPSCSGLVHLGVDRLYPOLAHNFFYSDRPREHFDVAFKSHRISDDPTIYLVAPEKTD 364

300 KMSNSLFVLYGLNHHHDLAAHTVCPGRRELIDEIFNHDGLASDFSILHAPCVTD 359

365 PAQAPAGCBIKILPHIPLDPPKLLTADYSALRERVLKERNGLTDLRQHIVTEYHW 424

360 SSLAPGCGSVYVLAIPVPHLGTANLDWTVEGPKLRDRIFAYLEQHYMPGLRSQIVTHRMF 419

425 TPLDIQAKYSNGQSYGVVADRFKNLGFAPQRSSLSNLYFVGGSVNPGGGMVMTLS 484

420 TPDPFDQNLAVHGSAPFVEPVLTOSAMFRPNRDKTITNLYLVGAGTHPGAGIPGVIGS 479

485 GOLVRDKIVADL 496

480 AKATAGLMLEDL 491

RESULT 4

RTI_AGRAU

ID AC RTI_AGRAU STANDARD; PRT; 501 AA.
PS4978;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44155;
RN SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
RT biosynthesis gene cluster and astaxanthin biosynthetic pathway
RT proposed at the gene level." (1995).
RL J. Bacteriol. 177:6575-6584 (1995).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
CC intermediaries of phytofluene, zeta-carotene and neurosporene by
CC the introduction of four double bonds (By similarity).
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
CC biosynthetic pathway.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D58420; BAA09594.1; -
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR008151; Phyt_n_dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phyt_n_dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SEQUENCE 501 AA; 54806 MW; 5P251AF11D679358 CRC64;

Query Match 24.8%; Score 645; DB 1; Length 501;
Best Local Similarity 30.8%; Pred. No. 1.2e-41;
Matches 151; Conservative 92; Mismatches 241; Indels 6; Gaps 2;

QY 10 IVTGAGLGLSAAISLATAGFSVQLIKNDKVGKLNIMTKDGTFDLGPSTLTMPHIFE 69
Db 12 IVTGAGFGGLAALRLQAGIPVLLQEKDKPGGRAYVNDQTRLEAQIQQFNPDRDVBGRQFL 71

QY 70 ALFTGAGKNMADYVQIQKVEPHWRNFFEDGSVIDLCEDAETORRELDKLGPGTYAQFOR 129
Db 72 ELWALTQGDARDVTLMPVSPYRLMWPQGVVYVNEADQLERQIAQFNPDDLEGYRRF 131

QY 130 LDYSKNLCTETAGYPAKGLDGFNDLLKFYGLPS---LISFDFVRSMDQGVRRFISDPKLVE 189
Db 132 RDYAEVYQD---GYVYKGLTVPFLKGLQMLKAAPAMKLEAYKSVKAVAFIKDPYLRQ 188

QY 190 ILNLYFTKYVGSSPYDAPALMNLPPYIQHYGLVYVKGMYGMAQAMEKLAVELGVEIRLD 249
Db 189 ARSYHTLLVGNPFSTSSIALNHALERENGWVFRGGTGALVQGMKILFQDLGGMLN 248

QY 250 AEVSEIQKQGRACAVKLANGDVLDPADIVVSNMEVTPAMEKLLRSPA---SELKKMQRPE 306
Db 249 AKVARIDTDPGRATGVTLADGRALTDAMVVASNGVWVHNRDRLTGARGQSPAKSLNAKR 308

QY 307 PSCSGLVHLGVDRLYPOLAHNFFYSDRPREHFDVAFKSHRISDDPTIYLVAPEKTDPA 366

```

b 309 WSNLSVFLHGLREAPKDVAAHTILFGPRYKELVNEIFKQPKLAEDPSLYLHSPCTTDP 368
y 367 QAPAGCEIIKILPHIPDLDPKLLTAEDYSALRERVVLKLERMLGLDLRQHVITYWTP 426
b 369 MAPPGQSTHYVLAPVHLGRADIDWAVEGPRYADRIILASLEERLIPLNLRANLTTTIFTP 428
y 427 LDIQAKYYSNQSIIYGVVADRFKILGKAPORSSELNLYFVGSVNPGGMPDWLTLSQ 486
c 429 SDFASLNAHGSARFSEIILTSQANFRPHNRDXTIRNFYLVGAGTHPGAGIFGVVGSAR 488
y 487 LVRDKIVADL 496
b 489 ATAQWMLSDL 498

RESULT 5
RTI_PHYBL
D CRTI_PHYBL STANDARD; PRT; 583 AA.
C P54982;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CARB.
S Phycomyces blakesleeanus.
C Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
C Phycomyces.
X NCBI_TaxID=4837;
P SEQUENCE FROM N.A.
C STRAIN=NRRL 1555;
A Ruiz-Hidalgo, M.J.;
L Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
C -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
C -!- COFACTOR: FAD (Probable).
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
C -----
C EMBL; X78434; CAAS5197.1; -
C PIR; S43139; S43139.
C InterPro; IPR002937; Amino oxidase.
C InterPro; IPR008150; Bac_phytoene_dh.
C InterPro; IPR000205; NAD_Bs.
C Pfam; PF01593; Phytn dehydro.
C ProDom; PD139017; Phytn dehydro; 1.
C PROSITE; PS00982; PHYTOENE_DH; 1.
C Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
C Transmembrane.
C NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
C TRANSMEM 531 551 POTENTIAL.
C SEQUENCE 583 AA; 65983 MW; 60E8F682B12FB591 CRC64;
Query Match 24.7%; Score 641.5; DB 1; Length 583;
Best Local Similarity 31.3%; Pred. No. 2.8e-41;
Matches 157; Conservative 100; Mismatches 223; Indels 21; Gaps 11;
2y 7 QRVIVIGAGLGLSAAISLATAGSVGLIKENKVGKLNIMTKGFTTDLGSLTTPH 66
3b 5 KHVIIIGAGGTATAAFLAREGKIVTVBKNFNGGRCSLINHGHRFDGQPSLYLMPK 64
2y 67 IFEALFTGAGKNADYYQIQKVEPHWENFFEDSGVLDLCEDAETQRELDKL-GPGTVAQ 125

```

```

Db 65 LPFEAFELDEKIEDHVELLRCHNNYKVHFDGDKIQSSLSRMKPMERIEGPD---G 121
Cy 126 FQRFELDYSKMLCTETBAQ---PAKGLDGFWDLLKF-YCPLRSLSLSPVFRSMQGVRRF 181
Db 122 FLRFDFMKESHYEGGVEMAIKONFETIWKILRLQVVP-ALFELHI FDFVYSRAAKY 179
Cy 182 ISPKLVEILNIYPIKVVSSPYDAPALANLLGLEYIQVHYGLA-VKGMKYGMAQAMEKLAVE 241
Db 180 FTKKMMMAFTPOSTMGMSPYDSPAVYNLLQTFABGIGVPRGFGNTVIOKLENIATE 239
Cy 242 -LGVEIRLDAVSEIQKD--GRACAVKLANGDVLPAIVV-INMEVIPAMEKILNSPASE 298
Db 240 KFGARFIYEAPVAKINTDDKGGKVTGVTLOSSEVTEADAVV-NADLVYAVHNLPPCRAW 299
Cy 299 LKKMQRFEPSCGLVHLGVDRLYLPOLAHENFFYSDHREHEDAVEKSHRLSDDDTIYLV 358
Db 300 TNYLAEKLLTSSSISFYWSLKRVPDELVDHNIPLAEAFKESFDEIFTDKWPSLSFYVN 359
Cy 359 APCKTDPAPAGACEIIKILPHIPHLDPDKLLTADYSAL-----RERVVLKLE-RMGULTD 413
Db 360 LPSRIDPTAAPGKDSMIVLVPIGHM-KSKTWEADYTWIVKARKMVLVLELRRLGILT 418
Cy 414 LRQHVITEEYVTPLDIOAKYYSNQSIIYGVVADRFKILGKAPORS--ELSNLYFVGS 471
Db 419 FIDLVEHEEVDPSIWKQKFNLRGSLILGLSHDVLQVLFPRPSTQDSTGRYKNLFFVGAS 478
Cy 472 VNPGGMPDWLTLSQGLVRDKI 492
Db 479 THPGTGVPIVLGSKLTSQV 499

RESULT 6
CRTI_MYXA
ID CRTI_MYXA STANDARD; PRT; 529 AA.
AC Q02861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CARC.
S Myxococcus xanthus.
C Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
C Cyctobacterineae; Myxococcaceae; Myxococcus.
X NCBI_TaxID=34;
P SEQUENCE FROM N.A.
C STRAIN=NR403;
C MEDLINE=93223667; PubMed=8467787;
C Fontes M., Ruiz-Vazquez R.M., Murillo F.J.;
C "Growth phase dependence of the activation of a bacterial gene for
C carotenoid synthesis by blue light.";
C EMBO J. 12:1265-1275(1993).
C -!- FUNCTION: This enzyme converts phytoene into lycopene via the
intermediaries of phytofluene, zeta-carotene and neurosporene by
C the introduction of four double bonds (by similarity).
C -!- COFACTOR: FAD (Probable).
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- INDUCTION: By blue light.
C -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
C -----
C EMBL; M94727; AAA25390.1; -
C PIR; S35306; S35306.
C InterPro; IPR000759; Adrndx reductase.
C InterPro; IPR002937; Amino Oxidase.
C InterPro; IPR008150; Bac_phytoene_dh.

```

RESULT 8

SEQUENCE FROM N.A.-

C STRAIN=ATCC 18366;

X MEDLINE=94368091; PubMed=8085820;
X Epubahead M Daub M E :
A

ENTENHALL M., DAUB M.B.: "Isolation, sequence, and character

T nicotianae phytoene dehydrogenase

L Appl. Environ. Microbiol. 60:2766

Db	228	PRGGPHKVLDAVXIGERMVGKYLANTGVSQVLTDGKKNGKPKATGVQLENGEVLNADL	287
Qy	278	VYSNNEVIPAEMKILIRSPASBLK-----KMORFFSPSCGLVHLGVDRLYPQLAHRNFFYS	333
Db	288	VVNADLVVYNNLLPKIEIGIKKYKANLRKASCSSISFYWSLSGMAKELETHNIFLA	347
Qy	334	DHPREHDAVFKSHRLSDDDTYILVAPCKTDPQAAPAGCE-IKILP--HI-----PHLD	385
Db	348	BEYKESFPAIFERQALPDDPSFYIHVPSRVPDPAAPDRDAVALVFGVGLHLQNGQPELD	407
Qy	366	PKMLTADYDYGALRERYLVKLE-RMGLTDLRQHVITEEYH-TPLDITQAKYYSNQGSYGV	443
Db	408	WPTLV-----SKARAGVLATIQARTGLS--LSPLITEEIVNTPTWTETKFNLSKGAILGL	460
Qy	444	VADRFKNLGRKAPORSSELNLYFVGSGVNPGGMPMTLSQLVPRDKIVAD	495
Db	461	AHDFPNVLAFRPTAQMGNALFVGASTHPTGTGVFIVLAGAKITAEQILEE	512
RESULT 9			
CRTI_RHOCA			
ID	CRTI_RHOCA	STANDARD;	PRT; 524 AA.
AC	P17054;		
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).		
GN	CRTI.		
OS	Rhodobacter capsulatus (Rhodospseudomonas capsulata).		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;		
OC	Rhodobacteraceae; Rhodobacter.		
OX	NCBI_TaxID=1061;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=SB1003 / St Louis, and BEC404;		
RC	MEDLINE=893113663; PubMed=2747617;		
RX	Armstrong G.A., Alberti M., Leach F., Hearst J.E.;		
RT	"Nucleotide sequence, organization, and nature of the protein		
RT	products of the carotenoid biosynthesis gene cluster of Rhodobacter		
RT	capsulatus".		
RT	Mol. Gen. Genet. 216:254-268 (1989).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=89327279; PubMed=2546948;		
RP	Bartley G.E., Scolnik P.A.;		
RA	"Carotenoid biosynthesis in photosynthetic bacteria. Genetic		
RT	characterization of the Rhodobacter capsulatus Crti protein.";		
RT	J. Biol. Chem. 264:13109-13113 (1989).		
RL	[3]		
RP	ERRATUM.		
RP	Bartley G.E., Scolnik P.A.;		
RL	J. Biol. Chem. 264:18260-18260 (1989).		
RL	[4]		
RP	SIMILARITY TO CAROTENOID DESATURASES.		
RP	MEDLINE=90368827; PubMed=2144293;		
RA	Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;		
RA	"Carotenoid desaturases from Rhodobacter capsulatus and Neurospora		
RT	crassa are structurally and functionally conserved and contain		
RT	domains homologous to flavoprotein disulfide oxidoreductases.";		
RT	J. Biol. Chem. 265:16020-16024 (1990).		
CC	-I- FUNCTION: This enzyme converts phytoene into zeta-carotene via the		
CC	intermediary of phytofluene by the symmetrical introduction of two		
CC	double bonds at the C-11 and C-11' positions of phytoene.		
CC	-I- Cofactor: FAD (probable).		
CC	-I- PATHWAY: Carotenoid and chlorophyll biosynthesis.		
CC	-I- SIMILARITY: Belongs to the phytoene dehydrogenase family.		

CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		

10

Journal of Management Inquiry 26(4) 401-417

```

53  TWPHI EALPTGCKWADYVOI QKVEHWRNFEDQSVDLDCEDAQTORRELDKLGPGT 122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67  TVPQSURDLWTKCGDFDADVELKDIPDFYVRWPGDSHFTVQGSTEAMKAEVARLSPGD 126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123 YAOFORFLBYSNMCLTETEAQYF-----AKGLDGFWDLLKF-----YGPLRSLLSDFVR 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 VAGYEXFL-----KQSEKRWFGYEDLGRSSMHKLDGLIKVLPFGMWEA-----DR 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
173 SMDQGVRRFISDPKLVEILNYFIKVYSSPYDA PALMNLIPYIQYHYGLWVKGYMGMA 232
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
174 TWYQHAALRVKDERLRMALSFHFLPTIGDDPNVTSWILVSQLEKSPGVHYAIGGVAALA 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
233 QAMEKLAVELGVEILRDAPSVSIOQQGRACAVKLANDVLPADIVVUSNMVEVLPAMEKLL 292
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 AAMAKVIECGGSGFRNMYTEVDILVEKGTATGVRLASGEVLRLAGLVSNADAGHTVNRLL 293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
293 RSP-----ASELKGMRPEFSCGIVLHL-----GVDRLYLPQLAHNFYFSDSHPREHFDAV 343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 RNHPRRRWTDAAVKSRRWS--MGLFVWYFGTKETKGMWDVGHHTIVNAPRYKGLVEDI 350
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 PAKSHRLSDDPITYLVAPCKTDPAQAPACEIILKIPHPHLDP-----DKLLTRADYSALR 399
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
351 FLKGLAKAKMSLYIHRPSITDPTVAPEGDDTFYALSPVPHLKQAQPYDQMOVAEPY---R 407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 ERLVYKLRMGLTDLRQHIVTBEVWTLDDIOAKYYSNQSGSYGVWADRFNQLGFKAPQBS 459
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 ESVLVLEQ-SNPGTGEAIGSLVPTETPFDRLYSPAGAGFSIBERILQSAWFRPHNIS 466
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
460 SELSNLYFVGGSVNPGGGMWMTLSGQLVRDKIVAD 495
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
467 EEVANILFLVGAGTHPGAGVPGVIGSAE-VMAKLAPD 501
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

SULT 11

CRTD_RHOSH STANDARD; PRT; 495 AA.

Q01671;

01-JUL-1993 (Rel. 26, Created)

16-OCT-2001 (Rel. 40, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)

Methoxyneurosporene dehydrogenase (EC 1.14.99.-).

CRTD.

Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.

NCBI_TaxID=1063;

[1]

SEQUENCE FROM N.A.

STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
MEDLINE=922307398; PubMed=1612412;
Gari E., Toledo J.C., Gibert I., Barbe J.;
"Nucleotide sequence of the methoxyneurosporene dehydrogenase gene
from Rhodobacter sphaeroides: comparison with other bacterial
carotenoid dehydrogenases";
FEMS Microbiol. Lett. 72:103-108(1992).

[2]

SEQUENCE FROM N.A.

STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
MEDLINE=95238278; PubMed=7721699;
Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;
"Complete DNA sequence, specific Tn5 insertion map, and gene
assignment of the carotenoid biosynthesis pathway of Rhodobacter
sphaeroides";
J. Bacteriol. 177:2064-2073(1995).

-1- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR
METHOXNEUROSPORENE TO SPHEROIDENE.

-1- COFACTOR: FAD (Probable).

-1- PATHWAY: Carotenoid and chlorophyll biosynthesis.

-1- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X63204; CAA44886.1; ..
DR EMBL; AJ010302; CAB38743.1; ALT_SEQ.
DR PIR; S23633; S23633.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR009150; Sac phytoene dh.
DR InterPro; IPR000205; NAD BS.
DR Pfam; PF01593; Amino oxidase; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP BIND 9 42 FAD (ADP PART) (POTENTIAL).
FT CONFLICT 115 115 A -> G (IN REF. 2).
FT CONFLICT 157 157 T -> P (IN REF. 2).
FT CONFLICT 170 170 L -> M (IN REF. 2).
FT CONFLICT 273 273 L -> I (IN REF. 2).
FT CONFLICT 440 443 PHGA -> ATGP (IN REF. 1).
FT SEQUENCE 495 AA; 52900 MW; 62227931415B253E CRC64;
SQ

Query Match 20.88; Score 541.5; DB 1; Length 495;
Best Local Similarity 29.98; Pred.No.9.6e-34;
Matches 151; Conservative 95; Mismatches 222; Indels 37; Gaps 13;

QY 8 RVIVIGAGLGGLSAAISLATAGSVQLIEKNKDVGGKINIMFKGFTFDLGFSLINMPhi 67
DB 7 KVVVVGGAGMGLSAIRLARAGCEVTLLRARAPGGRVRLTSPVAGPVDAGTPTVTLREV 66
QY 68 FEALPTGAGKNMADYVQIKVPEHWNFTEDGSVIDLCEDAETQRELDKL-GRGTYAQF 126
DB 67 FDDIFVCGQKLDHLLTLPQPLLAHWMWLDGSTLDLTTDLANVEAAVAFAGAREAPAP 126
QY 127 QRFLDYSKNLTCTEAGYFAKGLDGF-----WDLIKF-YGPIRLSLLSFDVF--RS 173
DB 127 RRFHDLNRL-----YDAFDRPMWEARFPLRAITATGALKAPRTWPALLPGMT 174
QY 174 MDQGVRFISDPKLVEILNFIKYKSSPYDAPALMNLPIQIHYGIVYKYGGMYGMAQ 233
DB 175 LDRILRFPDRRLRLQFRGYATYGGTYPGAPGLIALI-WAAEARGVWALTEGGMHRLL 233
QY 234 AMEKLAVELGVEIRLDAEVSIEIQKQDGRACAVKLANGDVLPADIVVSNMEVIPAMEKLR 293
DB 234 ALARLADQGVRLRYGAPVAGILRQGRPTGVQLADGRTLPAHDHIVFNGDPAALLAGCLG 293
QY 294 SPASELKKMQRTEP-SCSGLVLHLGVDRILYPOLAHNFFYSDHPRKHFDVAFKSHRLSDD 352
DB 294 DGPQDQVDFEPIRHPRLSLSAWWSYAAASGPPVLVHNVFFPADPRREFGPI-AAQOMPED 352
QY 353 PTIYLVAPCKTDAQAPAGCEIYKILPHIPHLDPDKLTAEDYSALREVRVIVKLERMGLT 412
DB 353 ATLVCIAEDRSQ-QQLPDGPERFEIIMNGP--PGRPAPKPEDFAQCRSEKTPDLRLQFELT 408
QY 413 DLRQHIIVTEWYT-PLDIQAKYYSNGSTYGVGVADRPKNLGFKAPQRSELSNLYFVGG 471
DB 409 --FDVPGFETSLTAPSGFASLPFASGGSYIG-LSPHGALASLKRPLARTALPGLWLAGG 465
QY 472 VNPQGGMPMTVLSQGLVRDKIVADL 496
DB 456 AHPGAGVPMALSGRAARAILADL 490

```

RESULT 12
CRTI_STRGR
ID CRTI_STRGR STANDARD; PRT; 507 AA.
AC P54381; P72447;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phycene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI OR CRTI.

```


Streptomyces griseus.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_taxid=1911;
[1]
SEQUENCE FROM N.A.
STRAIN=JA3933;
MEDLINE=97074881; PubMed=8917308;
Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
"Activation and analysis of cryptic crt genes for carotenoid
biosynthesis from Streptomyces griseus";
Mol. Gen. Genet. 252:658-666(1996).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L37405; AAA91950.1; -;
EMBL; X95596; CAA64850.1; -;
InterPro; IPR000759; Adrxn reductase.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac phytoene dh.
InterPro; IPR000205; NAD BS.
InterPro; IPR008151; Phyt. dehydro.
Pfam; PF01593; Amino oxidase; 1.
PRINTS; PR00419; ADXRDTASE.
ProDom; PD139017; Phyt. dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SEQUENCE 507 AA; 54509 MW; F8B977F7F569B2AC CRC64;

Query Match 20.3%; Score 526.5; DB 1; Length 507;
Best Local Similarity 28.6%; Pred No. 1.4e-32;
Matches 146; Conservative 88; Mismatches 218; Indels 59; Gaps 12;
9 VIVICAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGPSILTMPIHF 68
11 VVVVAGLAGLAAALHLGAGRRVTVVEREDVPGRAGLLESGGFRIDTGTPTVLTMPDLV 70
69 EALFTGAGKNWADYVQIOKVPHRNFFEDGSVIDLCEDAETORRELDKLGPGTYA--- 124
71 EDAFAAVERMADRRLRLAPAYARFADGSLDVTGDMGAARAAVEEFAGARQAVGYR 130
125 -----QFORFLDYSKNLCCTETEAGYFAGKLDGFWDLKFKYGP-LRSLSPDVF 171
131 RLRIWLERLYRVQMERFD-----TNFDSPLQLAHPDLARLAALGGF 172
172 RMDGGRVRFSDPKLVILYFYKYGVSSPYDAPALMULLPYQYHYGLVYVYKGMVGM 231
173 GRLDARIGHFVSDERLRVFSQALYAGVPARALAAAYAVIAYMDTVAGYVYFPGGMHAL 232
232 AQAMEKLAVELGVEIRLDAEYSEIQKQDGRACVAKLANGDVLPAIVVSNNEVIPAMEKL 291
233 PRAMADAASDAGAVLYRQVTRLEBSGDRVTAV-VTDQHI PCDAVVLTPD-LPVSRL 290
292 L-RSPASELXKMQREPCSGVLVLHLGVDRILYQLAHNFFYSDHPREHFDVFKSHRLS 350
291 LGRTFHRPLP-----LRHSPSAVILHTGTRDTPDLAHHHTISFGAAWNTFHELTTRGLM 346
351 DPTTYLVAPCKTDPQAQAPAGCEIILKILPHILD--PDKLLTAEDYSALRERVLVKLER 408
347 SDFSLLITPTADPSLAPPKHLHYVLAPCNTEVGPVREWRELGPVRELDLAELE 406

QY 409 MCLTDLRCHIVTEEVWTPLDIQAKYTSNQSGSIYGV--VADRFKNLGFKAQORSSEL--- 462
DB 407 RMPGELGAIAIEEGLVTPVD-----WTAQGHAAAGTSPFVAHTFFQTG--PFRPNLVRG 458
QY 463 -SNLYFVGSGVNPGGMPMTLSQLVRDKI 492
DB 459 TVNAVLAGCGTTPGVGVPTVLISGLAAERI 489

RESULT 13

CRTI_STRSE STANDARD; PRT; 508 AA.
AC P54971;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Streptomyces setonii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=38315;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISP 5395;
RA Hoshi K.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D55723; BAA09537.1; -;
DR InterPro; IPR000759; Adrxn reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR000205; NAD BS.
DR InterPro; IPR008151; Phyt. dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR ProDom; PD139017; Phyt. dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 508 AA; 54610 MW; 4B6DEFC076D51CB5 CRC64;

Query Match 19.8%; Score 514.5; DB 1; Length 508;
Best Local Similarity 28.0%; Pred No. 1.1e-31;
Matches 143; Conservative 89; Mismatches 221; Indels 57; Gaps 12;

QY 9 VIVICAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGPSILTMPIHF 68
DB 11 VVVVAGLAGLAAALHLGAGRRVTVVEREDVPGRAGLLESGGFRIDTGTPTVLTMPDLV 70
QY 69 EALFTGAGKNWADYVQIOKVPHRNFFEDGSVIDLCEDAETORRELDKLGPGTYA--- 124
DB 71 EDAFAAVERMADRRLRLAPAYARFADGSLDVTGDMGAARAAVEEFAGARQAVGYR 130
QY 125 -----QFORFLDYSKNLCCTETEAGYFAGKLDGFWDLKFKYGP-LRSLSPDVF 172
DB 131 RLRIWLERLYRVQMERFD-----TNFDSPLQLAHPDLARLAALGGF 173

173 SYDQVRRFTSDPKLVLEIINFIKYVSSPYDAPALMNLPIYQVHYGLWVYKGMGMA 232
 174 RLDRIGHVSDRLRRVFSQALYAGVTPARALAYAVIADTVAGVYPRGGHAPL 233
 233 QAMEKLAVELGVEIRLDAEVSRIQKDGACAVKLANGVLPADIVSNMEVIPAMEKLL 292
 234 RAMADAAADAGASFYQSVTLERSGDRVAV-VTDQERACDAVLTTPD-LPVSYRL 291
 293 -RSPASELKMQRPESCSGLVILHGVRLYPLAHNFFYSDBREHDFDAVFKSHRLSD 351
 292 GRSPHRLP-----LRHSPSAVILHAGTDTWPNLAHTISFGAWKSTFHELTRIGELMS 347
 352 DPTIYLVAPCKTDPAAQAPAGCEIIKILPHIPLD--PKLLTAEDYSALRERVLVKLRM 409
 348 DPSLLITRPASDPSLAPPSKHLHYVLAPCPNTEVPGVREWLCPRYDELLAELERR 407
 410 GLTDLRQHIWTEYWTPLDIQAKYVNSQSIYGV---VADRPNKLGFKAPORSSEL----- 462
 408 EMPGLGSAIEVEGLVTPVD-----WTAQGHAAAGTPPSVAHTFPQTG---PFRGNLVRGT 459
 463 SNEYFYVGGSVNPGGGMVMTLSQLVRDKI 492
 460 VNAVLAGCGTTPGCVPTVLISGLAAQRI 489

RESULT 14
 RTD RHOCA
 D CRTD RHOCA STANDARD; PRT; 494 AA.
 P17059;
 01-AUG-1990 (Rel. 15, Created)
 01-AUG-1990 (Rel. 15, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Methoxynurosporene dehydrogenase (EC 1.14.99.-).
 CRTD.
 S Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
 Rhodobacteraceae; Rhodobacter.
 NCBI_TaxID=1061;
 [1]
 N SEQUENCE FROM N.A.
 P STRAIN=SB1003 / St Louis, and BEC404;
 MEDLINE=89313663; PubMed=2747617;
 Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
 "Nucleotide sequence, organization, and nature of the protein
 products of the carotenoid biosynthesis gene cluster of Rhodobacter
 capsulatus.";
 Mol. Gen. Genet. 216:254-268 (1989).
 [2]
 SIMILARITY TO CAROTENOID DESATURASES.
 MEDLINE=90368827; PubMed=2144293;
 Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
 "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
 crassa are structurally and functionally conserved and contain
 domains homologous to flavoprotein disulfide oxidoreductases.";
 J. Biol. Chem. 265:16020-16024 (1990).
 1- FUNCTION: CONVERTS HYDROXYNUROSPORENE TO DEMETHYLSPHEROIDENE OR
 METHOXYNUROSPORENE TO SPHEROIDENE.
 1- COFACTOR: FAD (Probable).
 1- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 1- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL; X52291; CAA36537.1; --
 EMBL; Z11165; CAA77544.1; --
 PIR; S04406; S04406.
 InterPro; IPR002937; Amino_oxidase.

DR InterPro; IPR008150; Bac_phytoene_dh.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino Oxidase; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR ProSite; PS00982; PHYTOENE_DH; 1.
 KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 KW Oxidoreductase; FAD; Flavoprotein; NAD.
 FT NP BIND 8 41 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 494 AA; 52312 MW; D1180A023FEB5A9 CRC64;
 Query Match 18.9%; Score 490.5; DB 1; Length 494;
 Best Local Similarity 27.4%; Pred. No. 7.5e-30;
 Matches 141; Conservative 88; Mismatches 226; Indels 59; Gaps 14;
 QY 9 VIVIGAGLGLSAAISLATAGFSVQLIEKDKVGGKLNIMTKDGTDFDGLPILTMWHIF 68
 DB 7 VVVGARMGGLAAAI GAAAGLRVTVEAGDAPGKARAVPTFGGPADTGTVLTMHVL 66
 QY 69 EALPTGAGKNMADYVQIQKVEPHNRNFFEDGSGVIDLCDAETORRELDKL-GPGTYAQFQ 127
 DB 67 DALFAACGTAAEBHLTLIPLRLARHFWPDGSSLDLFTDTTEANIEAIRAFAGDKAAR 126
 QY 128 RFLDYKNLCTETEAGYFAKGLDGFWDLL-----KFGYPLRSLSPDVFRSMQGVTR 180
 DB 127 RFDLITLGLWEAFHRSVIAAPKPSDLWRIAAATVTRPQLMPALRPGL-----TMRDL 180
 QY 181 FISDPKLVLINFIKYVSSPYDAPALMNLPIYQVHYGLWVYKGMGMAEKLAV 240
 DB 181 HFKDPRLAQLFGRYATYVGRPGATPAVLISLWQAEVQ-GVAIRREGMEGVAAALARVAE 239
 QY 241 ELGVETRLDAEVSIEIQKDGACAVKLANGVLPADIVVSNMEVIPAMEKLLRSPA-SEL 299
 DB 240 AKGVRFHYGKAKRIVRKSGRTAVETIGVSIPGACIFNGDPGALRDGLLGDAARSM 299
 QY 300 KMQRPSPCSGLVILHGVRLYPLAHNFFYSDBREHDFDAVFKSHLSDDPTIYLV 359
 DB 300 EKSPRPAPSLSAWVAFGATPIGVDLAHENVFTADPELEFGPI-GAGNPREPIYI-- 356
 QY 360 PKCTD-PAQAPAGCEIIKILPHIPLDPKLIT-----AEDYSALRERVLVKLRM 409
 DB 357 -CAQDREMQAP-----VPEIERFEIIMNGPAGHQPPQEEAQCRAKTFPMLAAM 404
 QY 410 GLT---DLRQHIWTEYWTPLDIQAKYVNSQSIY-----GVVADRPNKLGFKAPORSSE 461
 DB 405 GLTFSPDPETRALT---TPALLSRFPGLSGIAYGGSPEGLTAT-----PRLPLARTG 454
 QY 462 LSNLYFVGGSVNPGGGMVMTLSQLVRDKIVAD 495
 DB 455 LKGLYLAGGTHFGAGVPAVLTSGTEAARALLAD 488

RESULT 15
 CRTI SYNY3
 ID CRTI SYNY3 STANDARD; PRT; 472 AA.
 AC P23273;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 GN PDS OR CRTD OR SIR1254.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=92256820; PubMed=1581575;
 Martinez-Perez I.M., Vioque A.;
 "Nucleotide sequence of the phytoene desaturase gene from
 Synechocystis sp. PCC 6803 and characterization of a new mutation
 which confers resistance to the herbicide norflurazon.";
 Plant Mol. Biol. 18:981-983 (1992).
 [2]
 REVISIONS TO C-TERMINUS.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:34:14 ; Search time 13.8338 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947A-24

Irfect score: 2598

Sequence: 1 MNSNDNQRIIVIGAGLGGLS.....MPMVTLSGQLVRDKIVADLQ 497

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

al number of hits satisfying chosen parameters: 283366

imum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287	49.5	497	E90061	hypothetical prote
2	779	30.0	517	T31463	probable diapophyt
3	773	29.8	517	S32169	hypothetical prote
4	770	29.6	492	A33120	phytoene dehydroge
5	685	26.4	492	S52586	phytoene dehydroge
6	672	25.9	502	B90061	squalene synthase
7	671	25.8	492	D37802	phytoene dehydroge
8	662.5	25.5	512	H84320	phytoene dehydroge
9	654	25.2	499	AQ2509	zeta-carotene desa
10	653	25.1	499	S43324	zeta-carotene desa
11	641.5	24.7	583	S43139	phytoene dehydroge
12	638	24.6	529	S35306	phytoene dehydroge
13	631.5	24.3	621	T48646	phytoene dehydroge
14	624.5	24.0	595	A35919	carotenoid biosynt
15	623	24.0	511	T50910	phytoene dehydroge
16	618.5	23.8	582	T46822	phytoene desaturas
17	595	22.9	448	B55546	crn protein - Sta
18	590.5	22.7	514	E69108	phytoene dehydroge
19	581.5	22.4	524	A32617	phytoene dehydroge
20	580	22.3	523	T36968	probable phytoene
21	568.5	21.9	536	B44327	phytoene dehydroge
22	554	21.3	490	AH1199	phytoene dehydroge
23	548	21.1	518	S39620	phytoene dehydroge
24	544	20.9	548	C75466	methoxynurosporen
25	543.5	20.9	495	S23633	methoxynurosporen
26	543	20.9	518	T50745	phytoene dehydroge
27	532.5	20.5	486	S49624	methoxynurosporen
28	521	20.1	525	T50893	methoxynurosporen
29	517.5	19.9	486	T50749	methoxynurosporen

30	516	19.9	488	2	AF1557	phytoene dehydroge
31	490.5	18.9	494	2	S04406	methoxynurosporen
32	465	17.9	528	2	T51119	phytoene desaturas
33	460	17.7	485	2	S32171	hydroxynurosporen
34	425.5	16.4	506	2	T34971	probable carotenoi
35	396	15.2	454	2	A9470	phytoene dehydroge
36	386	14.9	498	2	H83880	hypothetical prote
37	347	13.4	506	2	A2064	methoxynurosporen
38	322.5	12.4	490	2	D75297	methoxynurosporen
39	313	12.0	587	2	A86203	hypothetical prote
40	308	11.9	501	2	S75951	hypothetical prote
41	301.5	11.6	507	2	S74689	hypothetical prote
42	293.5	11.3	518	2	G90413	phytoene dehydroge
43	277	10.7	503	2	AC2446	hypothetical prote
44	252	9.7	511	2	R75561	probable phytoene
45	250.5	9.6	574	2	A96612	hypothetical prote

ALIGNMENTS

RESULT 1

E90061

hypothetical protein SA2351 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: E90061

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yurawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; PMID:21311952; PMID:11418146

A:Accession: E90061

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-497 <KUR>

A:Cross-references: GB:BA000018; PID:gl3702514; PIDN:BA43655.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA2351

Query Match 49.5%; Score 1287; DB 2; Length 497;

Best Local Similarity 51.9%; Pred. No. 3.8e-92;

Matches 254; Conservative 84; Mismatches 149; Indels 2; Gaps 2;

QY	7	QEVIVIGAGLGLSAAISATAGSVQLIENDKVGGKLNMTKDGFTDGLGSPSILTMPH	66
DB	3	KHIIVIGGGLGISAIRMAQSGSVSYLYEONTHIGKWKWHEHSDGFGDLGPSILLTPY	62
QY	67	IFEALFTGAGKNMADYQIQKVEPHWRNPFDDGSGVIDLCEBAETQRRELDKLGFGTYAQF	126
DB	63	IFEKLFYKSKQMSDYVTIKSLPHQWRSFPDGTIDLYEGIKETGCHNAILSQDISEL	122
QY	127	QRFLDYSKNLCTEIEAGYFAKGLDGFWDLLAFYGLRSLLSFDVFRSMDQGVRRFISPK	186
DB	123	QNYLNTARRIDRITKGYFYNGDLTSLQIIFHGLPLNALINDYVHTMQQADIKRISNPY	182
QY	187	LVYIILNYFTKYVGGSPYDAPALMMLLPYIQHYGLWYKGYGMAQAMEKLAVELGVEI	246
DB	183	LEQMLGYRIKYVGGSSSYDAPAVLSMLPHMQOQGLWYVEGGIHHLANALEKLAREGVTI	242
QY	247	RLDREVSHIQDGRACAVKLANGDVLPADIVSNMEVIPAMEKLRSFASLKKMOR-F	305
DB	243	HYGASVDNIKTYQRRVTVGRVLTGTGFEVKADYIISNMEVPTTYKYLIIHLDTORLNKLERF	302
QY	306	EPSCSGLVHLGVDRLYPQLAHNPFYSDHPREHFDVFKSHRSLSDDDPTIYLVPCKTDP	365
DB	303	EPASSGYMGLGVAQYQQLAHNPFETENAYLYQVFEKVLPPDDPTIYLVNTKTDH	362
QY	366	AQAPGCSIIKILPHIPLHDPDKLLTAEDYSALSERVLVKLERGLTDLRQHIYTEEWI	425
DB	363	TQAPGVYENIKVLPHIPIYIQ-DQPTTDEDYAKFRDKLDKLERMGLTDLRKHIYYSDVMT	421

128 RFLD--YSXNLCCTETEGYF-----AKGLDGFWDLLKTYGPIRLSLLSFVDFRSM---DOG 177
122 SFLTDVYK---YEIARRYFLERTYKPSD-----FYN-MTSLVQGGAKLTLNHAQOL 170
178 VRFPISDPKVEILNFIKVGSGSPYDAPALMNLPIQVHYGLWTVKGMYGMAQAMEK 237
171 IEHYIDNEIKQLLAFCTLYIGIDPKGSLYSIIIMWEMFGVHIKGMYGMAQGLAQ 230
238 LAVELGVEIRLDAVSE--IQKQGRACAVKLANGDVLPAIVWSNMEVIPAMEKLRLSP 295
231 LNKDLGWNIBLNAEIEGIIIDPKFKRADAIV--NGDIRKFCKILCTADFPFVAESLMPDF 289
296 ASELK-----XWQRPSPSCGLVHLGVD--RLYPQLAHNFFYSDDHREHEDAVEKSHRLS 350
290 APIKYPFKIADLDVSCAFLEWIGIDIDVTQVRLHNVISDDFRGNIEIFIG--RLS 348
351 DDPTIYLVAECTDPAPAGACEIIKILPHIPLD-----PKLLTAEDYSALRERV 403
349 YDSIYVYVPAVADKSLAPEGKTGIYVLPTEPKTGSGIDWSDEALTQO-----IKETIY 404
404 KLERNGL--TDLRQHIIVTEYWTPLDIAQKYNSQGSIVGVVADRPKNLGPAPQBSSEL 462
405 RKLATIEVFEDIKSHIVSEITFTPNDFPEQTHAKFGSAFGLMPTLAQSNYIRPQNVSRDY 464
463 SNLYFVGSGVNPFGGMPVMTLSQOLVRDKIVADLQ 497
465 KDLYFAGATHPGAGVPIVLTSAKITVDEMIKDIE 499

RESULT 7
37802
phytoene dehydrogenase (EC 1.3.-.-) crtI - Erwinia uredovora
Species: Erwinia uredovora
Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
Accession: D37802
Mitsawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim
i. Bacteriol. 172, 6704-6712, 1990
Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio
Reference number: A37802; MUID:91072214; PMID:2254247
Accession: D37802
Status: preliminary
Molecule type: DNA
Residues: 1-492 <MIS>
Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4127.1; PID:g216685
Superfamily: phytoene dehydrogenase
Keywords: oxidoreductase

Query Match 25.8%; Score 671; DB 2; Length 492;
Best Local Similarity 32.7%; Pred. No. 3e-44;
Matches 161; Conservative 89; Mismatches 230; Indels 12; Gaps 5;

11 VIGAGLGLSAAISLAPAGFSVOLIEKNDKVGKINIMTKDGTFFDLGSPSILTMPHIFA 70
6 VIGAGFGALALRLQAAGIPVLLLEQRDPGGRVYEDQGTFFDAGTPTVITDPSAIEE 65
71 LFTGAGKNADYVQIQVEPHWNPFRDGSVIDLCDAETQRELDKLGPGITVAOPREL 130
66 LFLAGKQKEYVELLPVTPFYRLCWESGVFNVDNDQTLKLEAQIQFNPDRVEGTREL 125
131 DYSKNLCCTEAGYFAGLGDGFDLLKXYGLRS---LLSFDVFRSMQGVRRFISDPKL 187
126 DYSAVFEK---GYLKLGTVPF---LSFDMRLAAQLAKLQAWRSVYSKVASIYEDHL 179
188 VEILNFIKVGSPYDAPALMNLPIQVHYGLWTVKGMYGMAQAMEKLAVELGVEIR 247
180 RQAFPSHLLVGGNPFATSIYTLIHALEREWGVNFRFGTGVQGMKLPQDLGGEVV 239
248 LDAVSEIQKQDGRACAVKLANGDVLPAIVWSNMEVIPAMEKL--RSPAS--ELKQMOR 304
240 LNAVSEVETTKNIEAVHEDGRRLFTQAVASNADVWHTYRDLLSQHPAAVKQSKLQT 299
305 PEPSCSGLVHLGVDRLYLPQLAHNFFYSDDHREHEDAVEKSHRLSDDPTIYLVAECKTD 364

300 KRMSNSLFIYFGLNHHHDLAHTVCFGRPRZELIDEIFNHDGLAEDFSLYLHAFCVTD 359
365 PAQAPACCEIIKILPHIPLDPPKLTASDYSALRERVLVKLERMGLTDLRQHIIVTEEYK 424
360 SSLAPEGCSYVYVLPVPLGTANLDMWTVEGPKLRDRIIFAYLEQHYMPGLRQSOLVTHRMF 419
425 TPLDIAQKYNSQGSIVGVVADRFKILGFKAPQBSSELNLYFVGSGVNPFGGMPVMTLS 484
420 TPDFRQLNAYGSAFSPVPLTQSAWFRPHNRDKTITNLVVGAGTHPGAGIPGVIGS 479
485 GQLVRDXIVADL 496
480 AKATAGLMLLEDL 491

RESULT 8
H84320
phytoene dehydrogenase [imported] - Halobacterium sp. NRC-1
Species: Halobacterium sp. NRC-1
CDate: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
Accession: H84320
RNg.W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, I.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; L
A>Title: Genome sequence of Halobacterium species NRC-1.
AReference number: A84160; MUID:20504483; PMID:11016950
Accession: H84320
Status: preliminary
Molecule type: DNA
Residues: 1-512 <STO>
Cross-references: GB:AE004437; NID:g10581152; PIDN:AAG19932.1; GSPDB:GN00138
Genetics:
Gene: crtII

Query Match 25.5%; Score 662.5; DB 2; Length 512;
Best Local Similarity 32.4%; Pred. No. 1.5e-43;
Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;

9 VIVIGAGLGLSAAISLAPAGFSVOLIEKNDKVGKINIMTKDGTFFDLGSPSILTMPHIF 69
9 VAVIGAGFGGLSTACYLADAGADVTVVEKTDQIGRASTLERDGRFDMGSPSYLMPDVF 68
69 BALFTGAGKNADYVQIQVEPHWNPFE-----DGSDVILCEDAET 110
69 EBFPAFSDSPSYDYGTLRLDPHYRIFFKDNEGRRFRDAPGLNVDTGDTIDVTDPREQ 128
111 QRRELDKLGPGTYAQFORFLDYSKNLCTETETAGYFAGLGDG-----WDLKFTY 160
129 VKQYFDAYEPGAGDVLDDYLAQAE-----NYEVGMEHFVKTRDPRVRDMDPKLAE 180
161 PLRSLLSFDVFRSMQGVRRFISDPKLVEILNFIKVGSGSPYDAPALMNLPIYIYHYG 220
181 YARGL---TLTGSMDQHEVQYFDPKLLQVQMOTLVFLGSPDPTTALYMLSHVDFGLG 237
221 LWYVKGSGVYMAQAMEKLAVELGVEIRLDAVSEIQKQGRACAVKLAN--GDLVPADIV 278
238 VTYPEGGGVWDGITHALRELGVFRTHPVSIAK---GRGCGFKLDTPEGDAVLADV 294
279 VSNMEVIPAMEKL----RSPASELKQWQRPSPSCGLVHLGVDRLYLPQLAHNFFYSD 334
295 VSDADYAHTEQALLPAQKQYDADYWDARTYAP--SAFLLYLGVGEGDVEPLAHTLVLS 352
335 HPRHFDVAPKSHRLSDDPTIYLVAECKTDPAQAPAG--CEIIKILPHIPLDPPKLTAE 393
353 DMDTHFAQIIFDDPAMPEDPAYILCVTSKTDVTVAPOGHNLPAVPAVAGLD--DTPAVRE 411
394 DYSALRERVLVKLERMGLTDLRQHIIVTEYWTPLDIAQKYNSQGSIVGVVADRPKNLGP 453
412 SY---RDLVDDIAENTGVLDRLRIIVVEERFSVSEPADRYNSHQGTALGLAHTLRQALP 468


```
Accession: S43139
Status: preliminary
Molecule type: DNA
Residues: 1-583 <RUI>
Cross-references: EMBL:X78434; NID:G468556; PIDN:CAA55197.1; PID:G468557
Introns: 197/2
Superfamily: phytoene dehydrogenase
Keywords: oxidoreductase

Query Match      24.7%; Score 641.5; DB 2; Length 583;
Best Local Similarity 31.3%; Pred. No. 7.6e-42;
Matches 157; Conservative 100; Mismatches 223; Indels 21; Gaps 11;

y 7 QRVIVGAGGGLSAAISLATAGFSVOLIEKNDKVGSKLNMTKDGFDTDLGSLTMDPH 66
b 5 KHWIIIGAGGGGATTAARLAREGIKTVVKEKNFGGRCRLNHNHGFDDQGPSLYLMPK 64

y 67 IFEALFTGAGKNMADYVOIQKVEPHWRNFFEDGSVIDLCEDAETORRELKLGPGTY 125
b 65 LFEAFEALEDEKIEDHVELLRCHNKKVHFDDGDKIQLSSDLRSKMPERIEGPD--G 121

y 126 FQRFIDYSKNLCTETAGY---PAKGLDGFMDLKP-VGSLRSLLSFDVFRSMDQVRRF 181
b 122 FLRFIDFMKESHYEGGVEMAIKQNFETIKWLRLQYVP--ALFLHIFDFVYRAAKY 179

y 182 ISDPKLVIELNYFIKYVSSPYDAPALMNLPIQVHYGLWYKGMYGMAQAMEKLAVE 241
b 180 PKTKGQEMAFTFQSMYKMSPYDSPAVYNLLQYTEPAEGIWKYFGGFTVIQKLENIATE 239

y 242 LGVEIRLDAVESEIQOD--GRACAVKLANGVLDADIVVSNMEVIPAMEKLRSASE 298
b 240 KFGARFYEAAPKADINTDRKGXVTGVLQSGVEIADAVCNADLVYAHNLLPFCRWT 299

y 299 LKKMORPEPCSGLVHLGVDRILYPOLAHNFFYSVDPREHFDVAFKSHRLSDPTIYLV 358
b 300 TNLAEKLTSSSISFVWSLKRVVPELDVNIFLAFAKESFDEIFTDHKMPSELSFYVN 359

y 359 APKTDPAQAPAGCEIILPHILPHLDPPKLLTAEDYSAL-----RERVLVKLE-RMGLTD 413
b 360 LPSSRIDPTAAPPGKDSMIVLPIGHM-KSKTNEADYTWIKRARKVLEVLRLGLTN 418

y 414 LRQHIVTEEYMTLDIOAKVYSNQSISYGVVADRFKNLGFKAPQRS--ELSNLYFVGS 471
b 419 FIDIVEHEEVNDPSIOWKKNLWRSILGSHLDVQLVWFRPSTQDSTGKYNKLPVGNAS 478

y 472 VNPGGGMPMTLSQLVRDKI 492
b 479 THPGTGVPIVLGSKLTSDQV 499

RESULT 12
33306
phytoene dehydrogenase (EC 1.3.-.-) - Myxococcus xanthus
A:Alternate names: phytoene desaturase
A:Species: Myxococcus xanthus
A:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 29-Sep-1999
A:Accession: S33306; S27594
A:Fontes: M.; Ruiz-Varquez, R.; Murillo, F.J.
A:Title: Growth phase dependence of the activation of a bacterial gene for carotenoid synthesis
A:Reference number: S33306; MUID:93223667; PMID:8467787
A:Accession: S33306
A:Molecule type: DNA
A:Residues: 1-529 <PCN>
A:Cross-references: EMBL:M94727; NID:G150079; PIDN:AAA25390.1; PID:G150080
A:Genetics:
A:Gene: CarC
A:Superfamily: phytoene dehydrogenase
A:Keywords: carotenoid biosynthesis; oxidoreductase

Query Match      24.6%; Score 638; DB 2; Length 529;
Best Local Similarity 30.3%; Pred. No. 1.2e-41;

Matches 155; Conservative 110; Mismatches 208; Indels 38; Gaps 10;

Qy 7 QRVIVGAGGGLSAAISLATAGFSVOLIEKNDKVGSKLNMTKDGFDTDLGSLTMDPH 63
Db 9 KHWIIIGAGGGGLSAAISLATAGGFRVTVVKEKAVPGKMKGLTLGNSGEYAVDTGPSILQ 68

Qy 64 MPHIFEALFTGAGKNMADYVOIQKVEPHWRNFFEDGSVIDLCEDAETORRELKLGPGTY 123
Db 69 LPGVLEQIFERRAARLEEDYVKLLPLDVNTRVHFMDGTDLDTTRHLDRMEAEALAKFGPQA 128

Qy 124 AQQRFLD-----YSKNLCTETAGYFAKGLDGFMDLKPFGP--LRSLSPDVF 172
Db 129 SAURQWDEGREKIGIAYOKFICTSA-----DNLGYAPWELAPTLRFXPKWQ 175

Qy 173 SMDQYRRFTISDPKLVIELNYFIKYVSSPYDAPALMNLPIQVHYGLWYKGMYGMA 232
Db 176 TLYRQLDGFDDRVTVYALAYPSKYLGLHPTTCSSVFSVIFLELAFGVHVEGGFREL 235

Qy 233 QAMEKLAHELVEIRLDAVESEIQODGRACAVKLANGVLDADIVVSNMEVIPAMEKL 292
Db 236 RGMWRCARDIGATFRGTPEKRVVDAGRAVGVLVGGEVLDADAVVWADLAYAARSLL 295

Qy 293 RSPASELK-----MORFEPSCSGLVHLGVDRILYPOLAHNFFYSVDHPRF-HFDVAFKS 346
Db 296 PAEABEGSLTDAALERAKEYSCSTFWAYVGLDVTYADLPHLIYLSESARRTDRDALEDR 355

Qy 347 HRLSDPTIYLVAPCKTDPQAOPAGCEIILPHILPHLD-PDKLLTAEDYSALREERVLVK 405
Db 356 HVDLEDPPFYCNPGVTPSPGAPAGESTLYLVPTNTGRPVQDWKTE--QALREIIPAM 413

Qy 406 LERMGTLDRLOHIVTEBYMTPLDIOAKVYSNQSISYGVVADRFKNLGFKAPQ-RSSELSN 464
Db 414 LEKVGLKGVREHREERYFPAETWRDDFNVFRAVFN-LSHTWLQGLPLRPVKYKRDIEG 472

Qy 465 LYFVGGSNPVGGMVMTLSGQLVRDKIVAD 495
Db 473 LYFVGGGTHPGSGLLTIMESAAADYLTRE 503

RESULT 13
148646
phytoene dehydrogenase (EC 1.-.-.-) [validated] - Cercospora nicotianae
A:Species: Cercospora nicotianae
A:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
A:Accession: T48646
A:Ehrnschaff, M.; Daub, M.E.
A:Appl. Environ. Microbiol. 60, 2766-2771, 1994
A:Title: Isolation, sequence, and characterization of the Cercospora nicotianae phytoene
A:Reference number: Z24498; MUID:94368091; PMID:8085820
A:Accession: T48646
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-621 <EHR>
A:Cross-references: EMBL:U03903; NID:G433144; PIDN:AAB86988.1; PID:G433145
A:Experimental source: ATCC 18366
A:Genetics:
A:Gene: PDH1
A:Function: <PDH1>
A:Description: EC 1.-.-.- [validated, MUID:94368091] phytoene dehydrogenase
A:Function: <PDH2>
A:Description: dehydrogenation of phytoene
A:Pathway: carotenoid biosynthesis
A:Superfamily: phytoene dehydrogenase
A:Keywords: oxidoreductase

Query Match      24.3%; Score 631.5; DB 2; Length 621;
Best Local Similarity 29.0%; Pred. No. 5e-41;
Matches 162; Conservative 105; Mismatches 214; Indels 77; Gaps 12;

Qy 10 IVIGAGLGLSAAISLATAGFSVOLIEKNDKVGSKLNMTKDGFDTDLGSLTMDPH 69
Db 11 IVIGVGGVSTAAARLARGFHVTVLEKKNPTGRCSLIHHEGTRFDQGPSLLLPGLFH 70
```


70 ALFTGAGKMM-ADYVQIQKVBPHWFFEDSGVIDLCEDAETORRELDKL-GPGTYAQFQ 127
71 RTFAELGSLQEGYKLLKCPNMYIHSEDEKTLSDLSVMKTEVKEGEGYTRYL 130
128 RFLQYSKNLCTETEAAGYFAKGLDGFMDLLKFGP--LRSLSPDFVFRSMOQGVRRFISDP 185
131 EPLKESHGHYELSVREVLRLNFEGLTAMLR---PEFLRHLLQLHPFFSIWTRACKYFWE 187
186 KLVEILNFIKYVGSSPDADALMNLFPYQVHYGLWTVKGMVGMQAQAKLAVELGVE 245
188 RLRRVFTTGSVMGMSPDAGTYSLLQYTLASGIWYVPGGFRVVEALVKIGERGV 247
246 IRLDAEVSII--QKQDGRACAVKLANGDVLDPADIVVSNMVEIPAMEKLL--RSPASELKK 301
248 FNETAVKXKLLSBDGVAQVELEDGRLEADVVVWNSDLVAYEXKLLPIKTPYAESLK 307
302 MQRFPSCGLVHLGVDRLYPOLAHNFFYSRPREHFDVAFKSHRLSDPTIYVAPC 361
308 GR--PGSCSSISFFWALDRQVPELBAHNFIADEYRHSFDSIFPKGHLIPDEPSFYVWPS 365
362 KTDPAQAPAGCEIILKILPHIL-----DPDKL-LT---A 392
366 RYDSTAAPGKDSVVVLVFGHLLBEDRHSQAQHSASRNHSSASPPDPQCLTPTXK 425
393 EYDSAL----RERVLVKLERGLTDLRQHIWTEYWTPLDIQAKYYSNQSGIYGVADRF 448
426 QWPMAMISLARKTILSTIQSRNTVDLTLIESTNSPLSNWKQTFNLDRAIGALGSHSFF 485
449 KNLGFKAPQRS-----SELSNLYFVFGSVNFGGG 477
486 NYLFRPTTRARKPCAFDQALLKFGVLGRAEVIIDAFGRGKDIKGLYVWGAHFGTG 545
478 MPWVLSQVLRDKIVAD 495
546 VPICLAGGALVAEQICGD 563

35919
35919
:Species: Neurospora crassa
:Date: 09-Nov-1990 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
:Accession: A35919
:Schmidhauser, T.J.; Lauter, F.R.; Russo, V.E.A.; Yanofsky, C.
:Title: Cloning, sequence, and photoregulation of al-1, a carotenoid biosynthetic gene
:Reference number: A35919; MUID:903771195; PMID:2144609
:Accession: A35919
:Status: preliminary
:Molecule type: DNA
:Residues: 1-595 <SCH>
:Cross-references: GB:M57465; GB:M33867; NID:G168746; PIDN:AAA33555.1; PID:G168747
:Note: the authors translated the codon CTC for residue 157 as Glu, and GAG for residue
:Superfamily: phytoene dehydrogenase
:Keywords: transmembrane protein

Query Match 24.0%; Score 624.5; DB 2; Length 595;
Best Local Similarity 29.7%; Pred No. 1.6e-40;
Matches 158; Conservative 110; Mismatches 201; Indels 53; Gaps 17;

5 DNQR---VIVIGAGLGSLAISLATAGFSVQLIEKNDKVGKLNIM-TKDGFTFDLGPS 60
3 ETQRPRAIIVGAGAGGIAVAARLAKAGVDVTLEKNDFTGRCGLIHTKAGYRFDQGS 62
61 ILTMPIHEALPTGAGKMM-ADYVQIQKVBPHWFFEDSGVIDLCEDAETORRELDKL- 118
63 LLLLPLGPRETFEDLGTLEQEDVELLQCFPNYINWSDGRKFPSTENATMKVEISKWE 122
119 GPGTYAQFQFIDYSKNLCTETEAAGYFAKGLD-----GFWDLKLFYGP-----LR 163
123 GPD---GFRVYL-----SWLAEGHQHYETSLSRHLVHRNFKSIILEADPLRVVT 167
164 SLLSFDVFRSMOQGVRRFISDPKILVEILNFIKYVGSSPDADALMNLFPYQVHYGLWY 223

Db 168 LLWALHPFESIWHRAGRYFKTDRMORVFTFATMTMGSPFPDAPATYSILQYSELAEGTWY 227
QY 224 VKGGMVYGAQAAMEKLAVELGVEIRLDAEVSIIQKDG-----RACAVKLANGDVLPA 277
Db 228 PRGGPHKVLDAVLKIGERMGVKYRLNTGVSQVLTGDKGNGKPKATGVQLENGEVLNADL 287
QY 278 VVSNNMEVIPAMEKLLRSPASELK-----QMRFPSCGLVHLGVDRLYPOLAHNFFYS 333
Db 288 VVVDNLVTVYNNLLPKBEIGGKKYANKLNKAKSCSSISFYWSLSGMAKELETHNIFLA 347
QY 334 DHPREHFDVAFKSHRLSDPTIYVAPCKTPDAPACAGE-IILKLP--HI-----PHLD 385
Db 348 EYKESFDPAIFERQALPDDPSFYIHVPSRVPDPSAAPPDRDAVIALVPVGHLLQNGQP 407
QY 386 PKLKTAEIDYSAERLRLVKLE-RMGLTDLRQHIWTEYWTPLDIQAKYYSNQSGIYGV 443
Db 408 WPTLV-----SKARAGVLATIOARTGLS--LSPLITEIVNTPTWETKENLSKAILGL 460
QY 444 VADRFKNLGFKAPQRSSELSNLYFVFGSVNFGGMVMTLSGOLVRDKIVAD 495
Db 461 AHDFFNLAFRPTTRARKQGMNAYFVGASTHEFTGTVPIVLAGAKITAEQILRE 512

RESULT 15
T50910
phytoene dehydrogenase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C;Accession: T50910
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynt
A;Reference number: Z25270
A;Accession: T50910
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-511 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BAA94063.1
A;Experimental source: strain III44
C;Genetics:
C;Superfamily: phytoene dehydrogenase

Query Match 24.0%; Score 623; DB 2; Length 511;
Best Local Similarity 31.1%; Pred No. 1.7e-40;
Matches 154; Conservative 106; Mismatches 215; Indels 20; Gaps 10;

10 IVIGAGLGSLAISLATAGFSVQLIEKNDKVGKLNIMTKDGFTFDLGPSILTMPIHE 69
22 LVVSGFGGMAAAVRLAAKGYRVTLEKLDAPGGRAYVHREGHVFDAGPTIVTVPLFD 81
70 ALFTGAGKMMADYVQIQKVBPHWFFEDSGVIDLCEDAETORRELDKLGPQTYAQF 129
82 ELWALGRKFSDDIEKSLDPFYRIRFDGDHFDYSGDPARMRAEVRIRISDAEGFERF 141
130 LDYSKNLCTETEAAGYFA---KGLDGFMDLLKFGPRLSLLSFDVFRSMOQGVRRFISDPK 186
142 M-READQC--YELGFTLGDKAFDVTGDLIK-AAFL--IVKLRGWRSLHQVSSHLKHPK 195
187 LIVEILNFIKYVGSSPDADALMNLFPYQVHYGLWTVKGMVGMQAQAKLAVELGVEI 246
196 LRIANSLQSLIGGNPFSVTSYALVNALERQWVHVMGTTGELIRGLVDVDFGGMGT 255
247 RLDAEVSIIQKQDGRACAVKLANGDVLDPADIVVSNMVEIPAMEKLL---RSPASELKK 302
256 RLKAEVKRIEVDNGVATGVTLADGERIPADIVVCGDGTGVLKYLKNEVDARWKKHWT-DAR 314
303 QRFEPSCGLVHLGVDRLYPOLAHNFFYSRPREHFDVAFKSHRLSDPTIYVAPCK 362
315 ERGHYSMGLFWYVFTDRRYEDVPHHMMVGLGPRRELDDIFRKKCLASDFSITLHRTA 374
363 TDPAQAPAGCEIILKILPHIL--DPDKLKTAEIDYSAERLRLVKLRLVGLTDLRQHI 420

b 375 TDPSPAGCDTFYALMPVHLSGTDWTTQAEPY---RQSVQEALERTVLPGLGHLRV 431
y 421 EBYWTHLDIOAKYYSNCGSIYGVVADRFYNLCFKAPORSSELSNLYFVGGSVNPQSGMEM 480
b 432 SFCTTFLDFQHRLLSYKGAGFGLLEPLLOSAYFRPHNRSEDEVKNLFMWGASTHPGAGVFG 491
y 481 VTLGGQLVRDKIVAD 495
b 492 VMSAKAL-ESVLPD 505

Search completed: February 29, 2004, 14:52:41
Job time : 15.8338 secs